

LLM Evaluation Report

Executive Summary

Methodology

The evaluation methodology for the WAISE project is designed to assess self-hosted Large Language Models (LLMs) within XWiki, focusing on their performance in knowledge management and technical support tasks.

- **Task Selection:** Tasks are chosen to cover essential areas such as content summarization, question answering, and other knowledge management functions.
- **Benchmarking Framework:** A benchmarking framework is established to measure LLM performance using automated metrics and manual evaluation.
- **Ethical Considerations:** The evaluation includes considerations for data privacy and the environmental impact of the models.
- **Iterative Improvement:** The methodology involves continuous refinement of models and evaluation metrics based on feedback.

Find the full methodology document at:

<https://design.xwiki.org/xwiki/bin/view/Proposal/X-AI/WAISE/Evaluation%20Methodology/>

Metrics

Answer Relevancy

Measures the quality of the RAG pipeline's generator by evaluating how relevant the actual output is compared to the provided input. Score is calculated as: Number of Relevant Statements / Total Number of Statements.

Faithfulness

Evaluates whether the actual output factually aligns with the contents of the retrieval context. Score is calculated as: Number of Truthful Claims / Total Number of Claims.

Contextual Precision

Measures the RAG pipeline's retriever by evaluating whether relevant nodes in the retrieval context are ranked higher than irrelevant ones.

Contextual Recall

Evaluates the extent to which the retrieval context aligns with the expected output.

Custom Context Relevancy

Assesses how well the retrieval context aligns with the information required to generate the expected output. It evaluates the proportion of information from the expected answer found within the retrieval context.

Note: All metrics are self-explaining LLM-Eval, providing reasons for their scores.

Evaluation Criteria:

Summarization: Alignment, Coverage
Text_generation: score
Rag-qa: AnswerRelevancy, Faithfulness, ContextualPrecision, ContextualRecall, CustomContextRelevancy

Model Information

Model	Context Length	Provider	License
GPT4o	128k	OpenAI	Proprietary
GPT4o-mini	128k	OpenAI	Proprietary
claude3_5_sonet	200k	OpenRouter	Proprietary
mistral2_large	128k	Mistral	Mistral Research License
llama3_1_402b	128k	OpenRouter	LLAMA 3.1
llama3_1_8b_Q4	128k	Ollama	LLAMA 3.1
mixtral-8x22b	16k	Mistral	Apache 2.0
mistral-nemo_12b_Q4	128k	Ollama	Apache 2.0
gemma2_9B_Q4	8k	Ollama	Gemma ToU
phi3_mini-128k_4b_Q4	128k	Ollama	MIT License
phi3_medium-128k_14b_Q4	128k	Ollama	MIT License
command-r_35B_Q4	128k	Ollama	CCA-NonCommercial 4.0
qwen2_7b_Q4	128k	Ollama	Apache 2.0

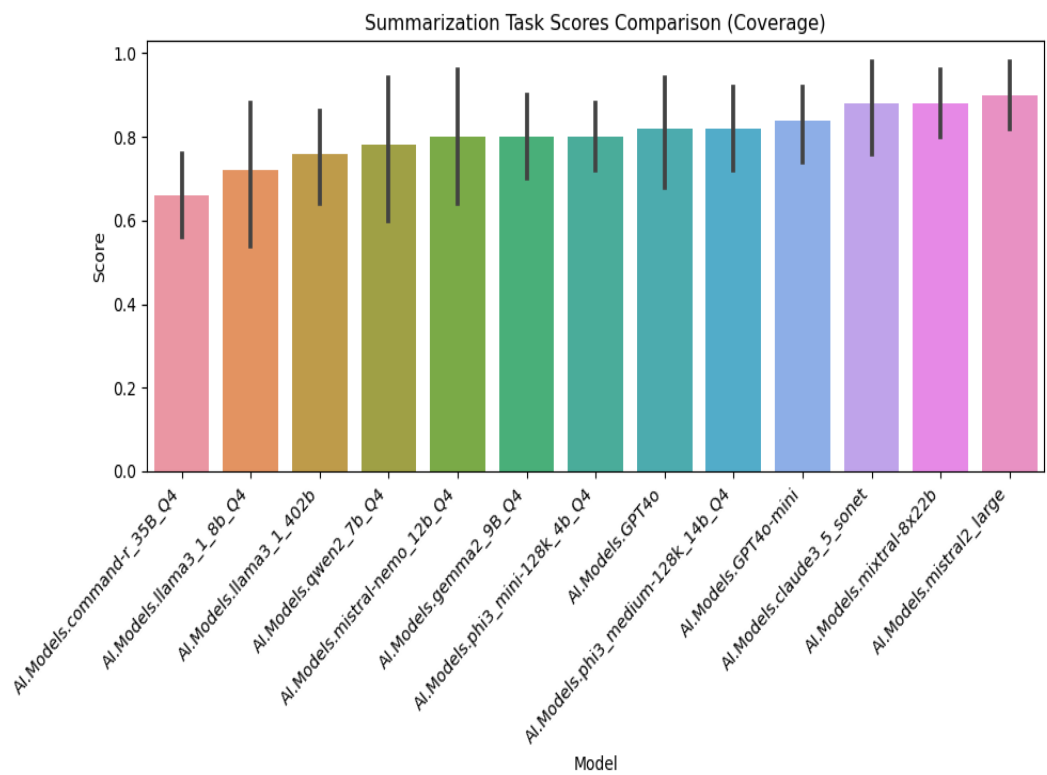
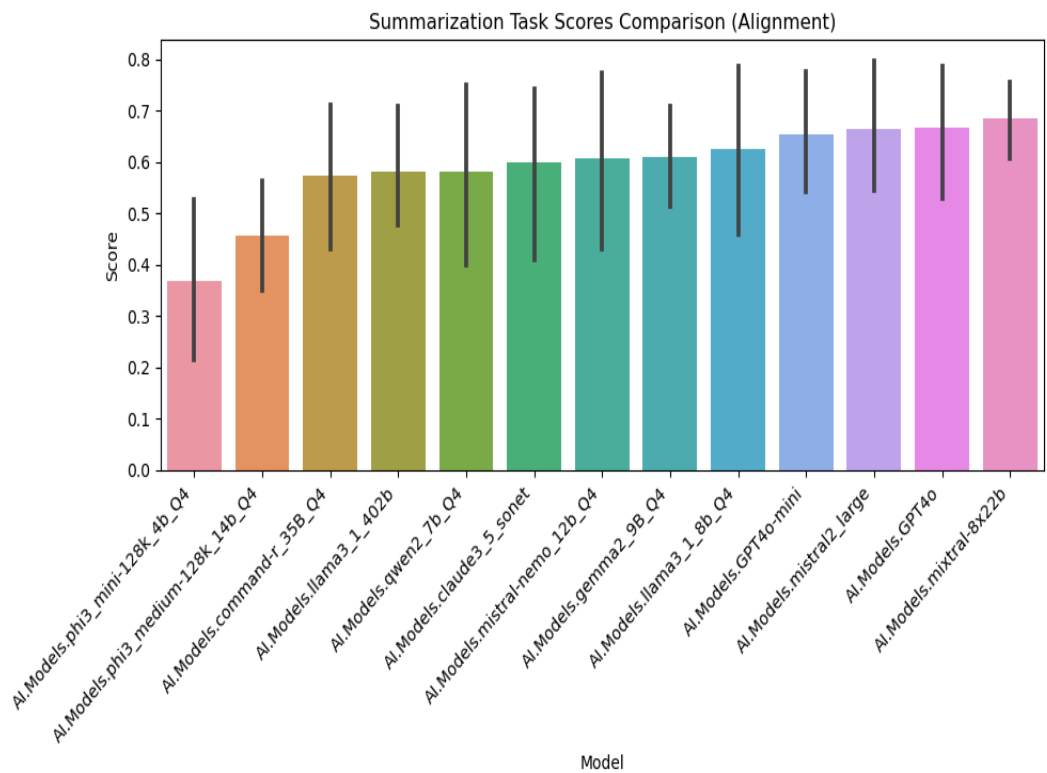
Setup Information

Task	Model	Temperature	Power Measurement
------	-------	-------------	-------------------

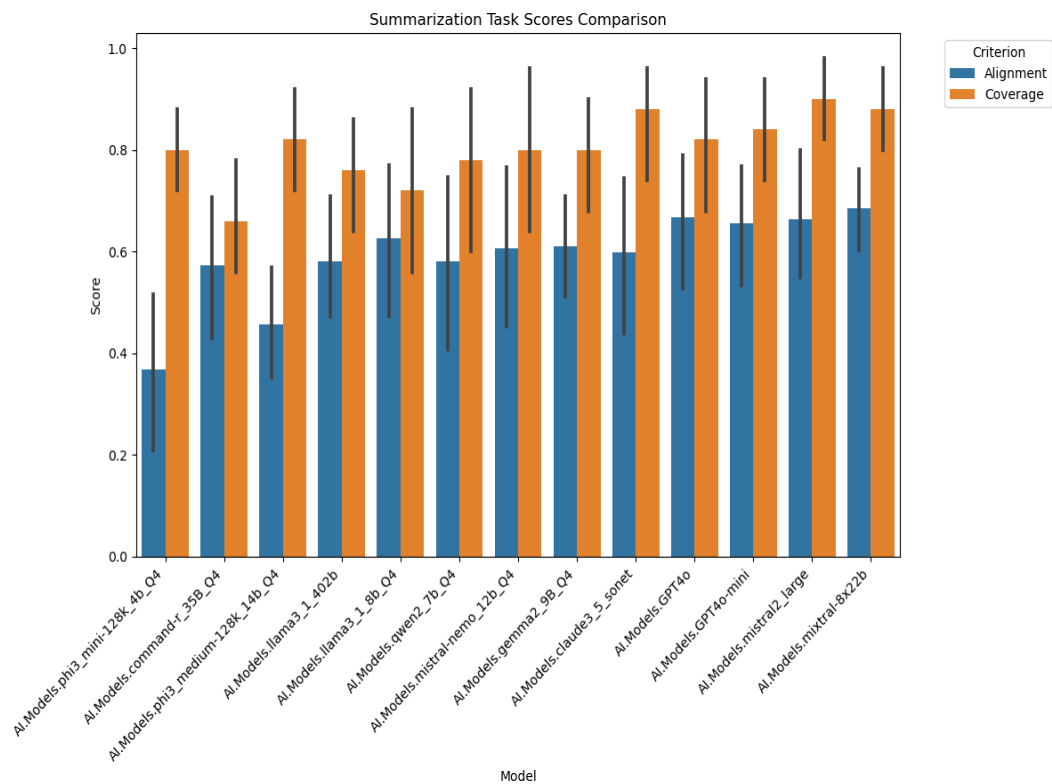
RAG-qa	AI.Models.qa_GPT4o	0.8	No
RAG-qa	AI.Models.qa_GPT4o-mini	0.8	No
RAG-qa	AI.Models.qa_claude3_5_sonnet	0.8	No
RAG-qa	AI.Models.qa_mistral2_large	0.8	No
RAG-qa	AI.Models.qa_llama3_1_402b	0.45	No
RAG-qa	AI.Models.qa_llama3_1_8b_Q4	0.45	Yes
RAG-qa	AI.Models.qa_mixtral-8x22b	0.8	No
RAG-qa	AI.Models.qa_mistral-nemo_12b_Q4	0.3	Yes
RAG-qa	AI.Models.qa_gemma2_9B_Q4	0.3	Yes
RAG-qa	AI.Models.qa_phi3_mini-128k_4b_Q4	0	Yes
RAG-qa	AI.Models.qa_phi3_medium-128k_14b_Q4	0.8	Yes
RAG-qa	AI.Models.qa_command-r_35B_Q4	0.3	Yes
RAG-qa	AI.Models.qa_qwen2_7b_Q4	0	Yes
summarization	AI.Models.GPT4o	0.3	No
summarization	AI.Models.GPT4o-mini	0.3	No
summarization	AI.Models.claude3_5_sonnet	0.3	No
summarization	AI.Models.mistral2_large	0.3	No
summarization	AI.Models.llama3_1_402b	0.3	No
summarization	AI.Models.llama3_1_8b_Q4	0.3	Yes
summarization	AI.Models.mixtral-8x22b	0.8	No
summarization	AI.Models.mistral-nemo_12b_Q4	0.3	Yes
summarization	AI.Models.gemma2_9B_Q4	0.3	Yes
summarization	AI.Models.phi3_mini-128k_4b_Q4	0.3	Yes
summarization	AI.Models.phi3_medium-128k_14b_Q4	0.8	Yes
summarization	AI.Models.command-r_35B_Q4	0.3	Yes
summarization	AI.Models.qwen2_7b_Q4	0.3	Yes

text_generation	AI.Models.GPT4o	0.3	No
text_generation	AI.Models.GPT4o-mini	0.8	No
text_generation	AI.Models.claude3_5_sonnet	0.8	No
text_generation	AI.Models.mistral2_large	0.8	No
text_generation	AI.Models.llama3_1_402b	0.8	No
text_generation	AI.Models.llama3_1_8b_Q4	0.8	Yes
text_generation	AI.Models.mixtral-8x22b	0.8	No
text_generation	AI.Models.mistral-nemo_12b_Q4	0.8	Yes
text_generation	AI.Models.gemma2_9B_Q4	0.8	Yes
text_generation	AI.Models.phi3_mini-128k_4b_Q4	0.8	Yes
text_generation	AI.Models.phi3_medium-128k_14b_Q4	0.8	Yes
text_generation	AI.Models.command-r_35B_Q4	0.8	Yes
text_generation	AI.Models.qwen2_7b_Q4	0.8	Yes

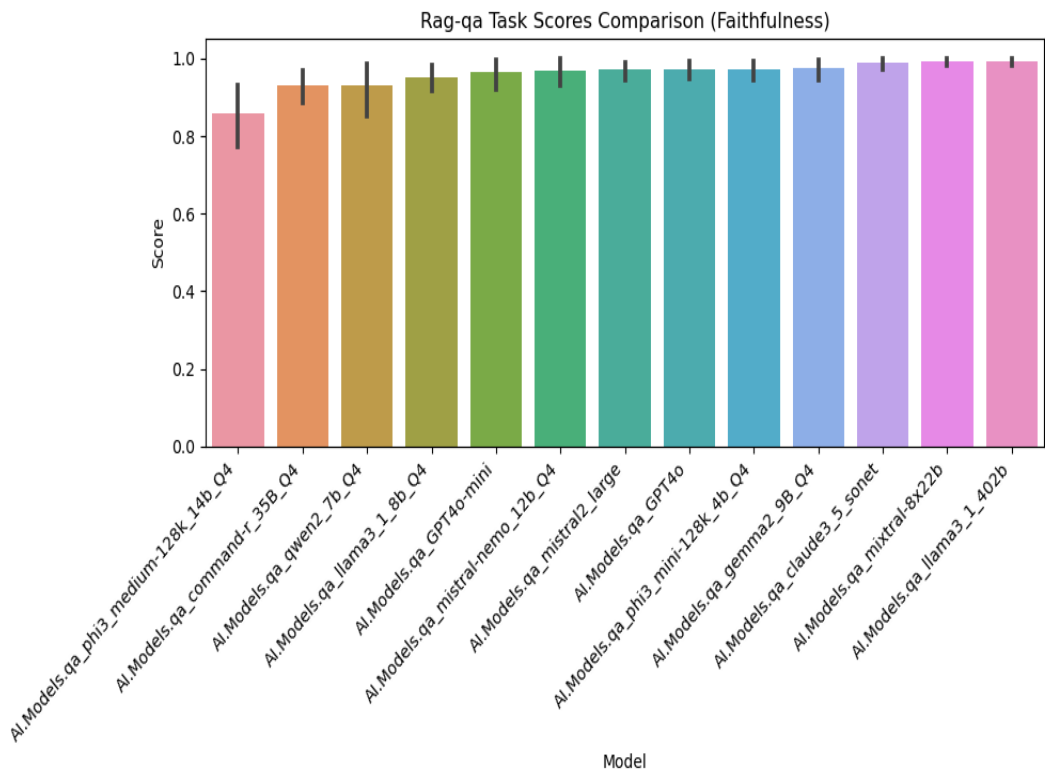
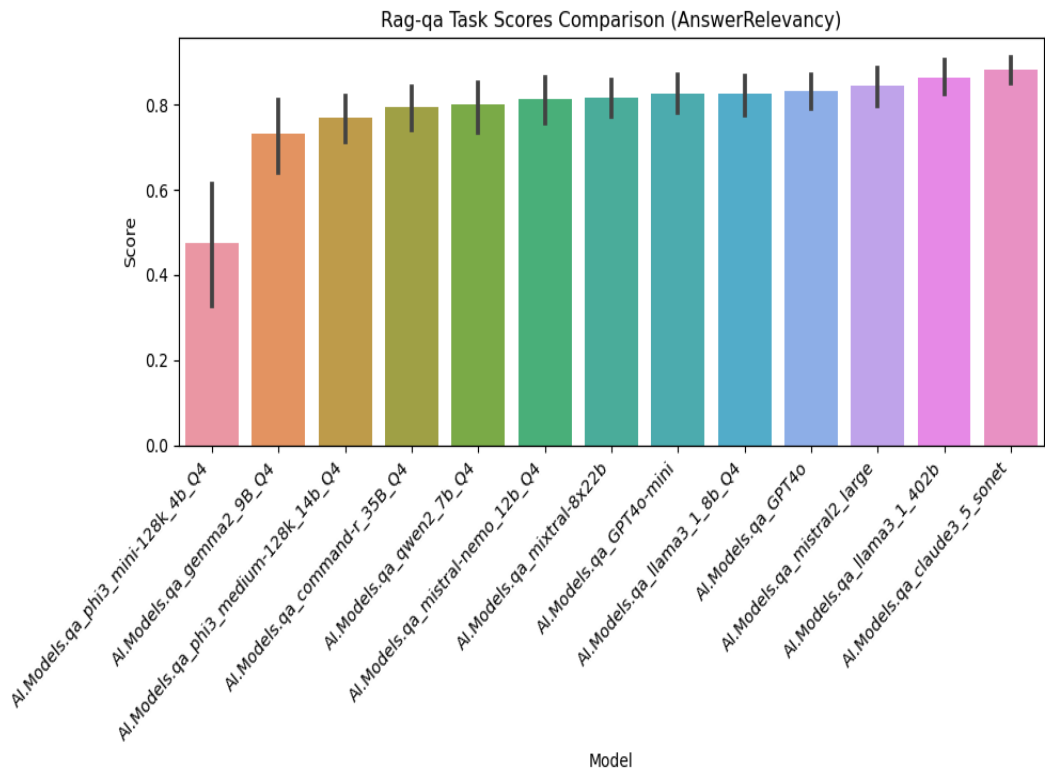
Summarization Task Results

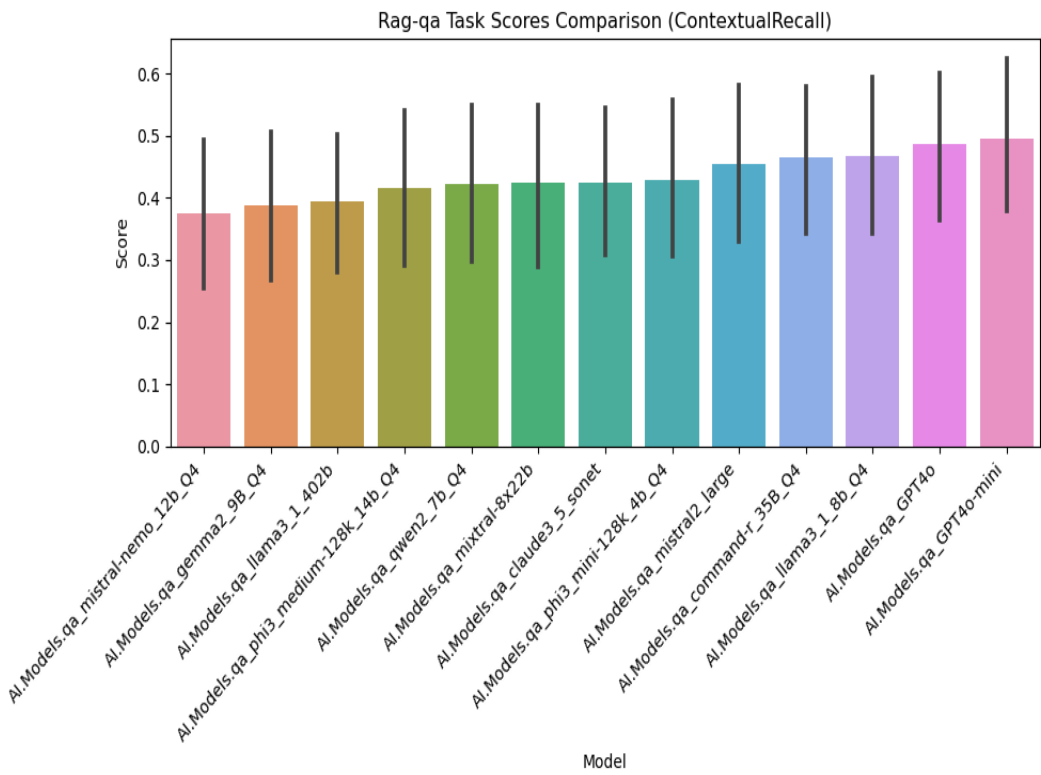
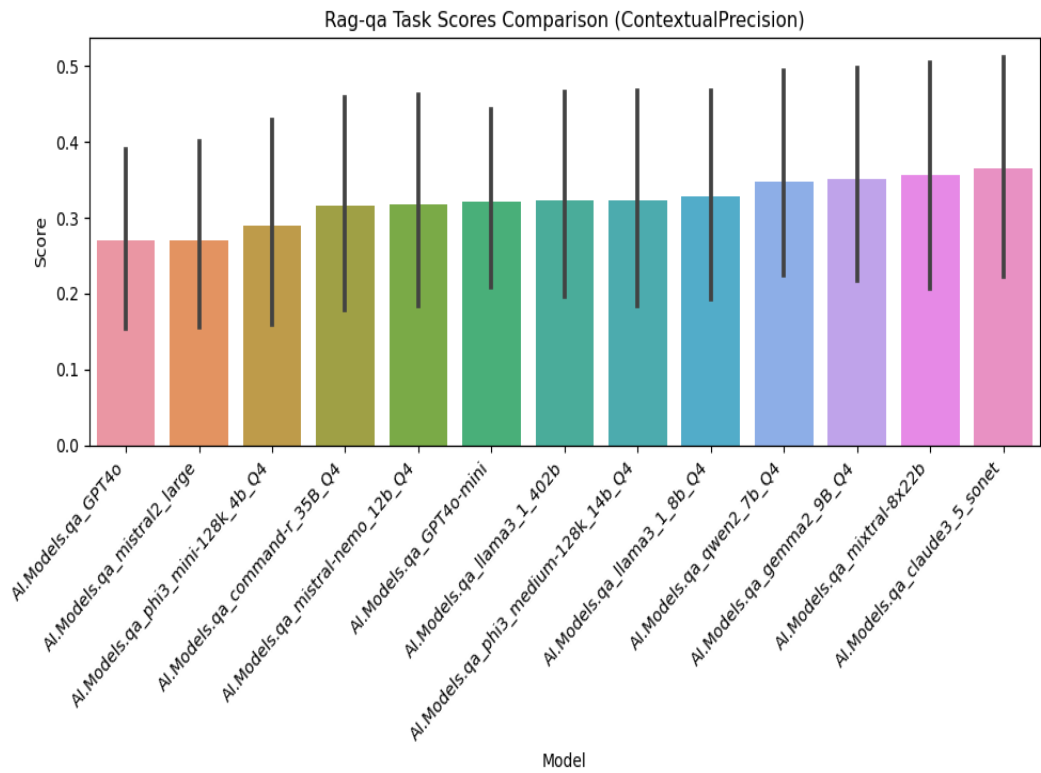


Grouped Results

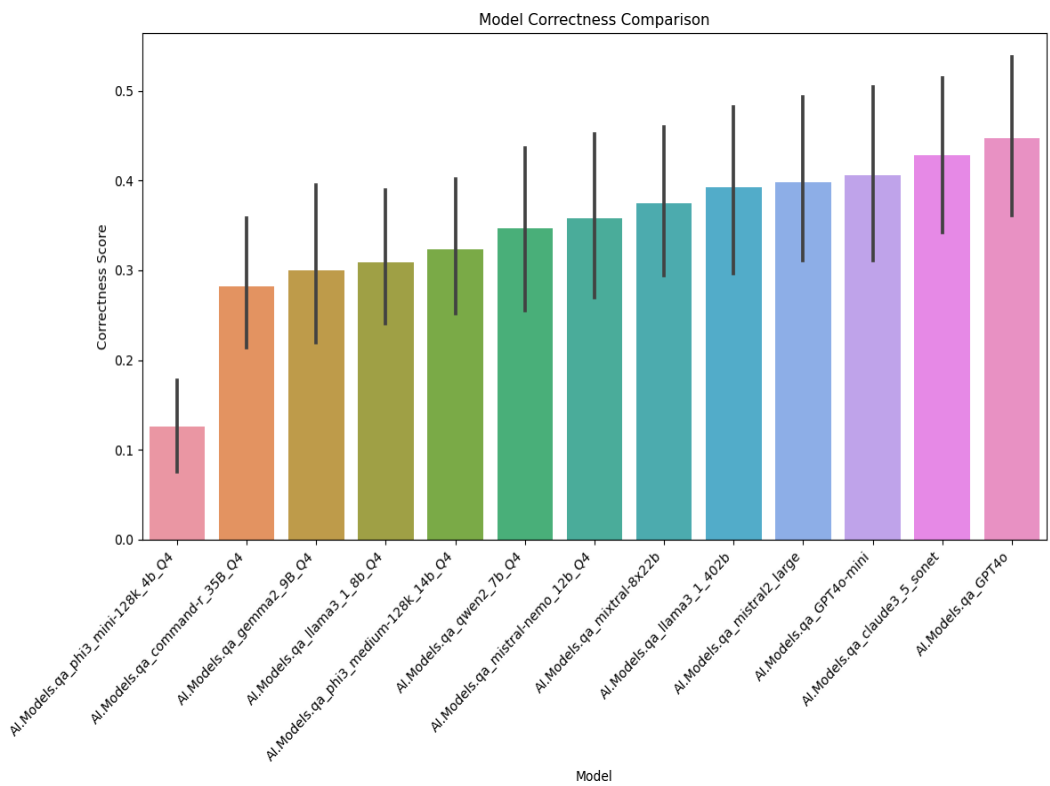


Rag-qa Task Results

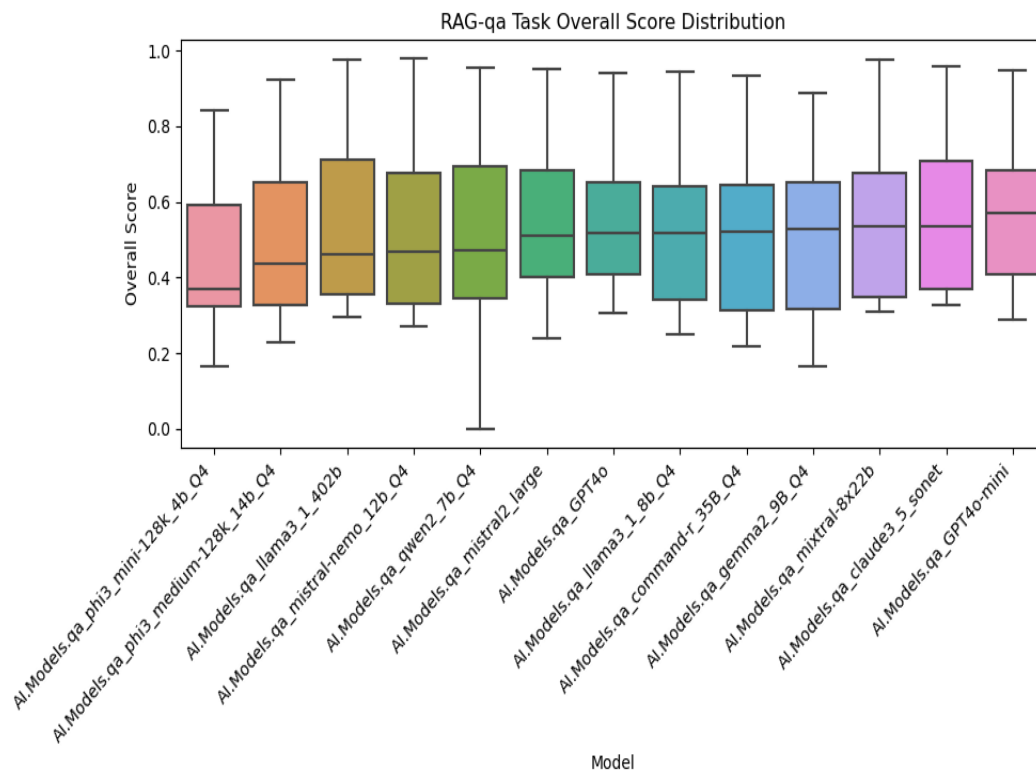




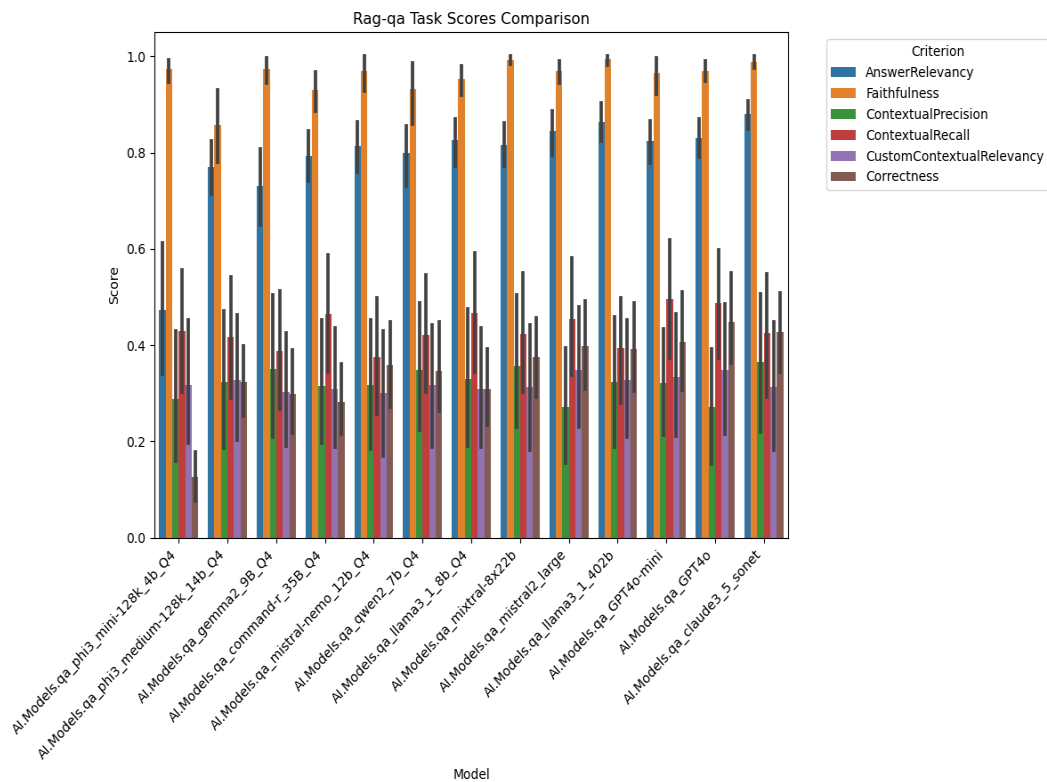
Model Correctness Comparison



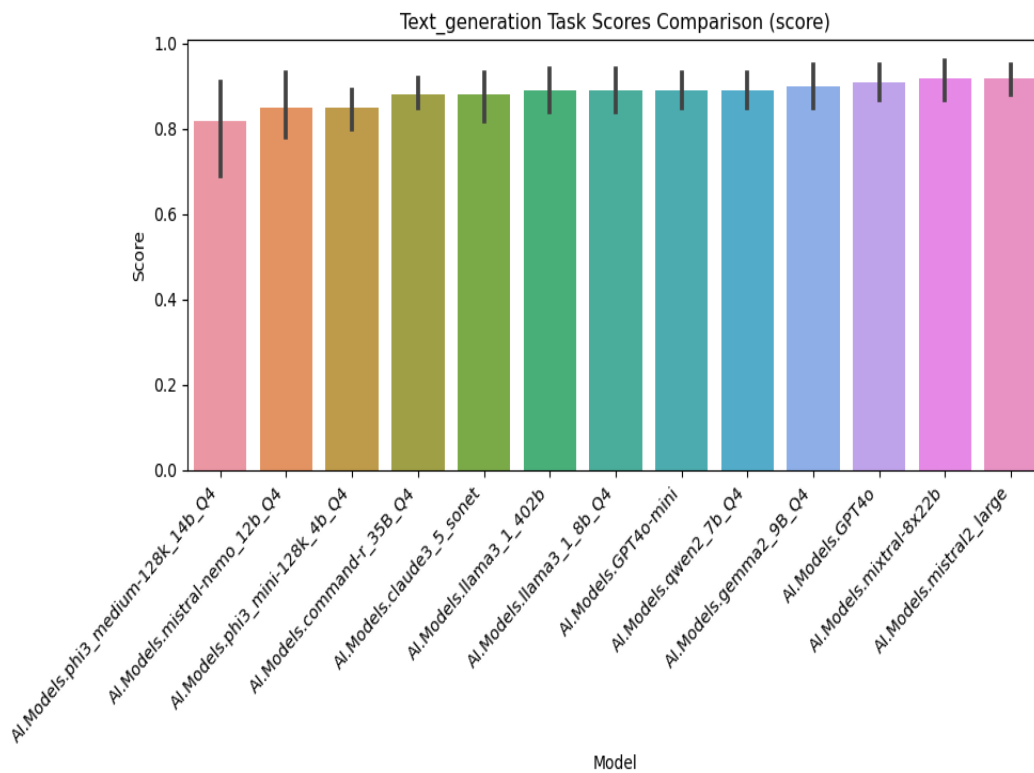
Overall Score Distribution



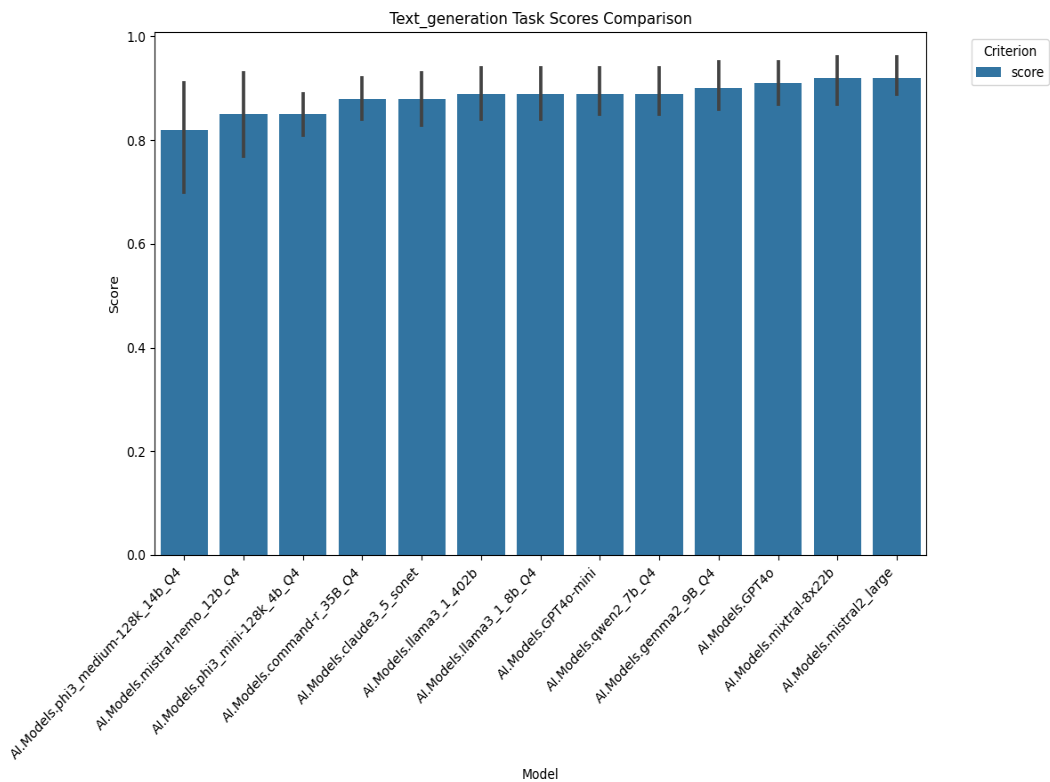
Grouped Results



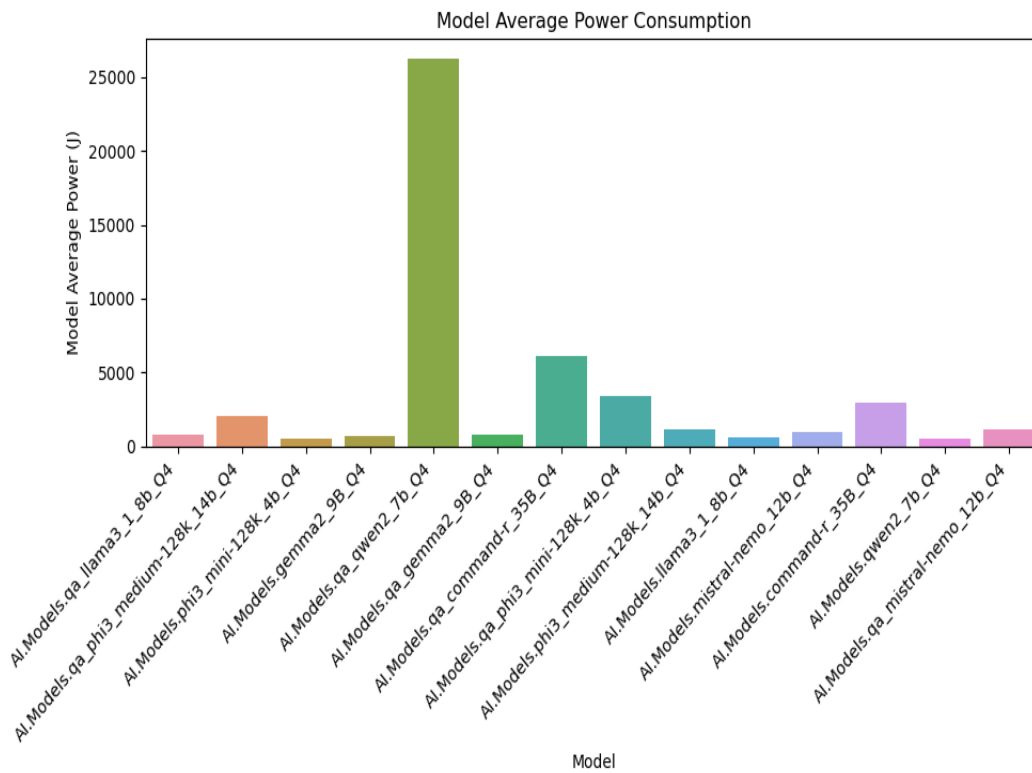
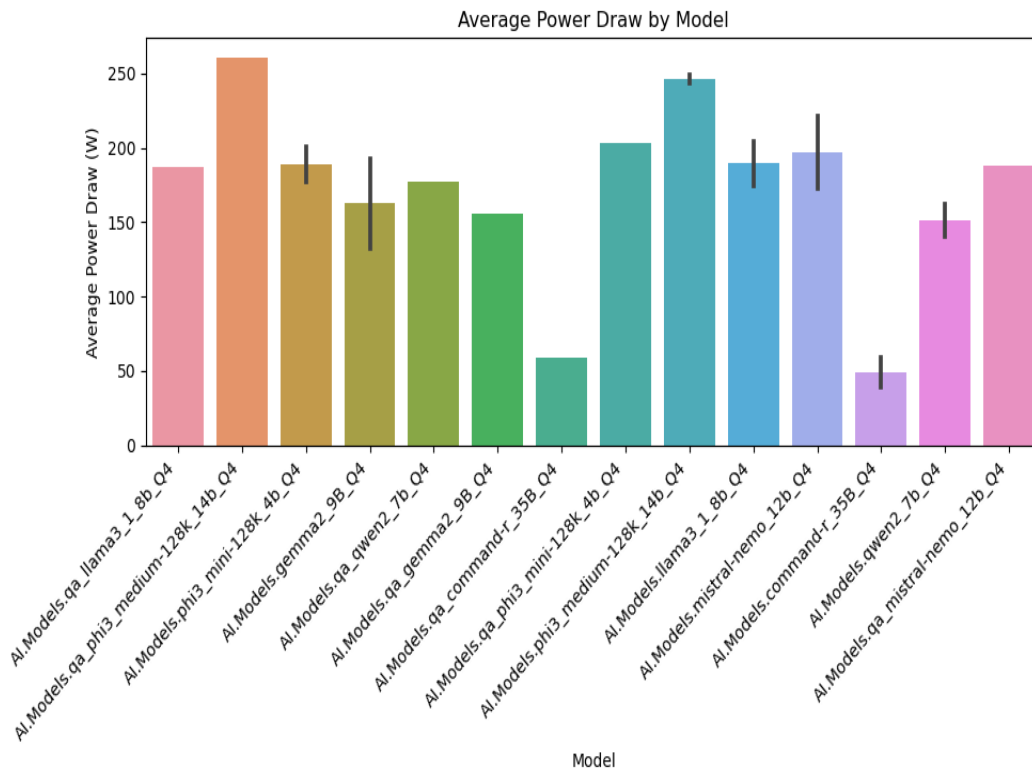
Text_generation Task Results

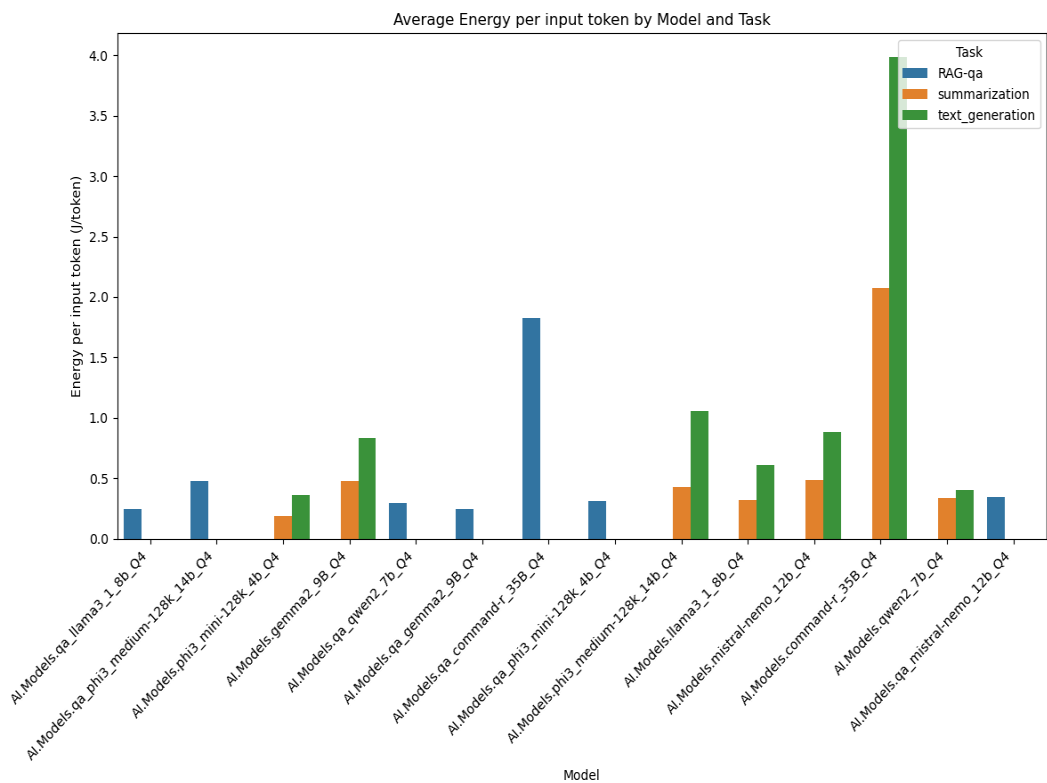
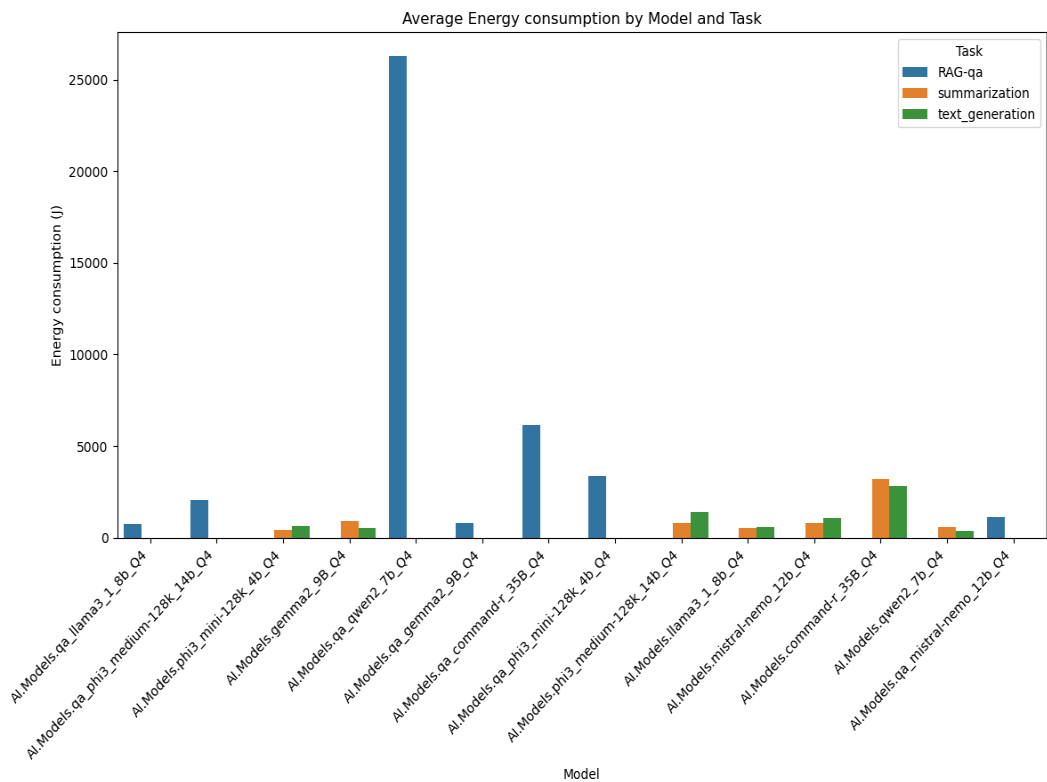


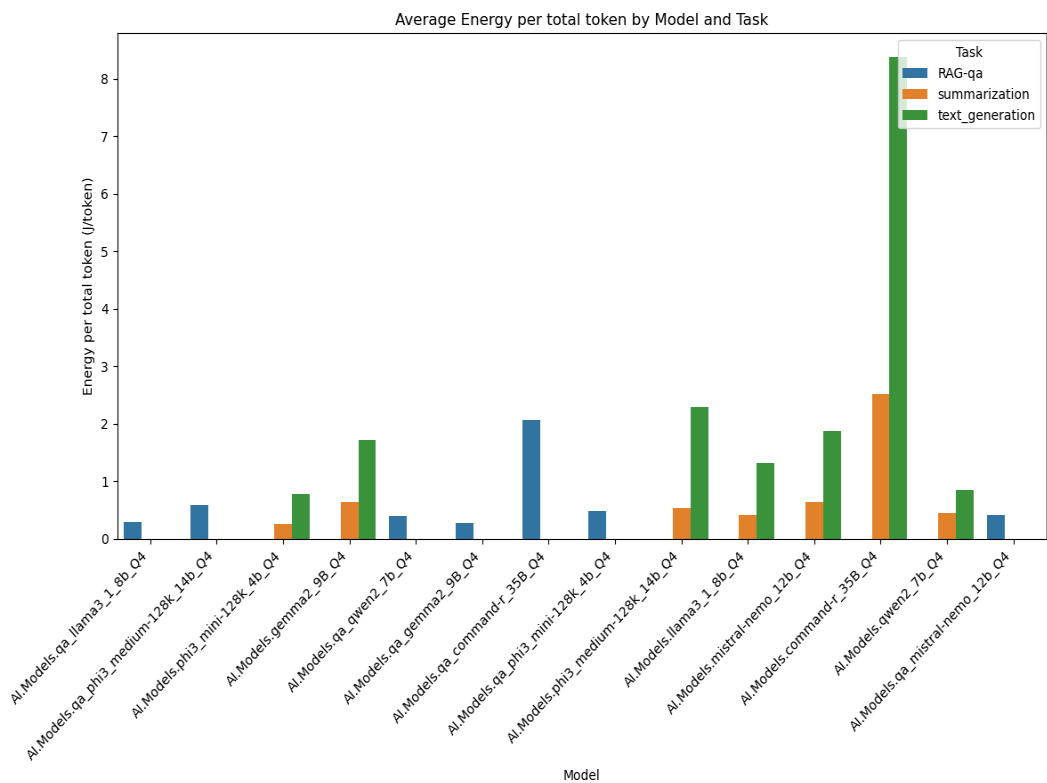
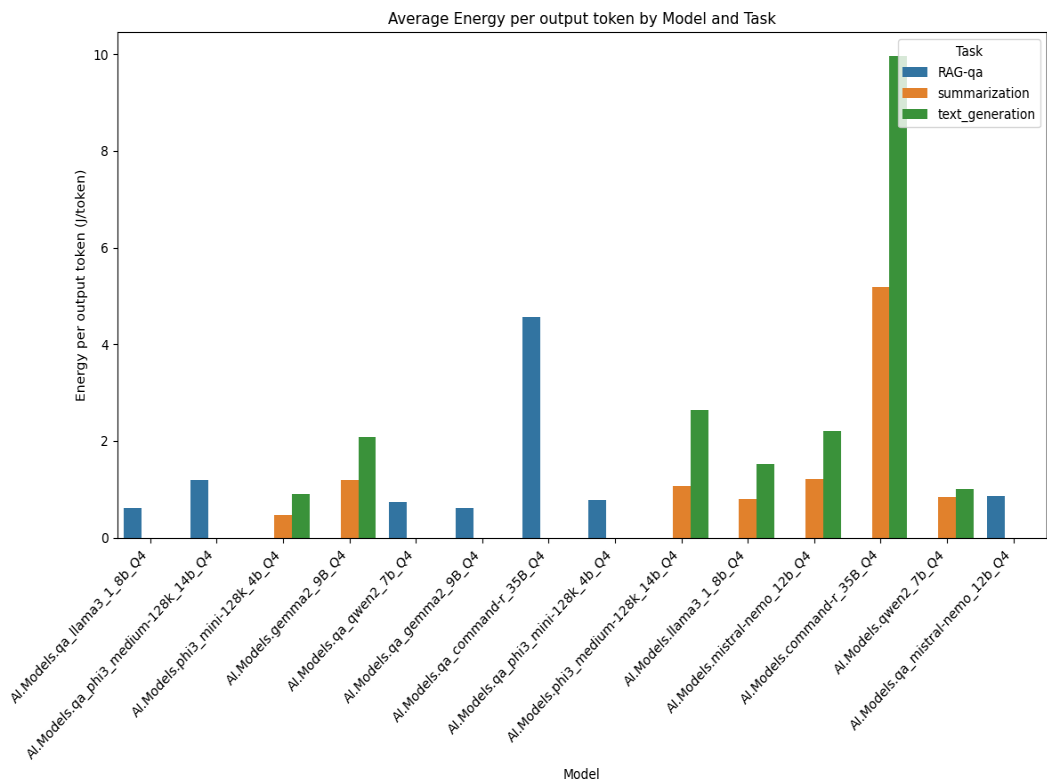
Grouped Results



Power Consumption Results







Evaluation Results

Rag-qa Results

Model: AI.Models.qa_GPT4o

File: qa_001_result.json

Overall_score: 0.39796888705758016

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 0.875

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.17947998901214762

Reasons:

AnswerRelevancy: The score is 0.83 because the response largely addresses the issue of the missing 'Bell' icon in XWiki, but includes several irrelevant source links and a general comment about dependencies that do not directly contribute to solving the problem.

Faithfulness: The score is 0.88 because the actual output incorrectly claims that the Extension Manager can manually install the 'Notifications Application' by uploading the XAR file, whereas the retrieval context clearly states that offline installation of extensions is not supported in XWiki.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'installing extensions using the Extension Manager' and does not mention enabling notifications or the `xwiki.properties` file, which is unrelated to the issue of the 'Bell' icon not being available. Similarly, the second node addresses 'issues related to administration access and configurable applications,' which also does not pertain to the problem at hand. None of the nodes provide relevant information on enabling notifications or fixing the bell icon issue, leading to a contextual precision score of 0.00.

ContextualRecall: The score is 0.50 because while the retrieval context mentions configuring notifications in the wiki, it lacks specific details about setting 'notifications.enabled' in 'xwiki.properties' to 'true'.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications through the `xwiki.properties` file or the `notifications.enabled` setting.

Correctness: The actual output discusses troubleshooting notifications but does not specifically mention enabling them via the `notifications.enabled` setting in `xwiki.properties`, which is the key detail in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.7107189196970228

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.7

ContextualRecall: 0.5714285714285714

CustomContextualRelevancy: 0.6

Correctness: 0.3928849467535657

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the issue of not receiving notifications in XWiki. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.70 because the first node in the retrieval context is relevant, explaining 'how to follow a user to receive notifications,' and is ranked higher than irrelevant nodes. However, the second node discusses 'filters for hidden pages and other technical details,' which do not directly relate to enabling notifications, and should be ranked lower. Additionally, the fifth node, which is relevant and 'provides information about the notifications settings menu,' is ranked lower than several irrelevant nodes, such as the third node about 'customizing email templates for notifications,' which does not directly help in resolving the issue.

ContextualRecall: The score is 0.57 because while the retrieval context nodes support several aspects of the expected output, such as watching pages or users and using the alert menu, they lack information on enabling notifications in 'xwiki.properties' and subscribing to pages or users, which are crucial details in the expected output.

CustomContextualRelevancy: The retrieval context covers enabling notifications via user profile and watching users, but lacks details on setting 'notifications.enabled' in 'xwiki.properties' and specific steps to watch pages or entire wikis.

Correctness: The actual output covers some aspects like watching users or pages and enabling notifications but includes many irrelevant points such as email settings, admin settings, and filtering options. It lacks direct instructions on setting 'notifications.enabled' to 'true' in 'xwiki.properties' and omits details about using the 'alert' menu.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9414102361952148

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.898461417171289

Reasons:

AnswerRelevancy: The score is 0.75 because the response addresses the main question about enabling notifications for one's own actions in XWiki, but includes multiple irrelevant source links that do not directly contribute to the solution.

Faithfulness: The score is 1.00 because there are no contradictions. Everything aligns perfectly with the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant nodes in the retrieval context are perfectly ranked higher than the irrelevant nodes. The first node states, 'By default, one don't receive notification about one's own activity... You have the ability to disable this filter in the "Advanced filtering options" section of your own notification settings,' directly addressing the input question. The second node also aligns with the input by mentioning, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter.' All irrelevant nodes, such as the third node discussing following a user and receiving notifications about events triggered by them, are correctly ranked lower.

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context contains all the information from the expected output, including the default behavior of not receiving notifications about one's own activity unless explicitly targeted and the ability to disable this filter in the 'Advanced filtering options' section of the user profile.

Correctness: The actual output accurately reflects the default notification settings and the ability to change them. It mentions the default filter and steps to disable it, aligning with expected output. The only minor discrepancy is the actual output's additional detail about the 'Own event filter' naming.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5954516738733368

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8333333333333334

ContextualPrecision: 0.33730158730158727

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Correctness: 0.3520751226051001

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about disabling unwanted notifications in XWiki without any irrelevant information. Great job!

Faithfulness: The score is 0.83 because the actual output incorrectly claims the existence of 'Advanced filtering options' in the notifications settings, which is not supported by the retrieval context.

ContextualPrecision: The score is 0.34 because the relevant nodes in the retrieval context are ranked lower than several irrelevant nodes. For example, the first node discusses 'following a user and receiving notifications about their activities,' which is not directly related to disabling notifications for pages, yet it is ranked higher than relevant nodes. Similarly, the second node focuses on 'selecting the level of detail in email notifications' and the third node is about 'enabling or disabling the email feature,' both of which do not address the input query. Relevant nodes, such as the fourth node that mentions 'configuring notification settings,' are ranked lower, affecting the score.

ContextualRecall: The score is 0.75 because the retrieval context supports the idea of automatic page watching and changing settings in user profiles, as mentioned in the 5th node in retrieval context, but it does not specifically mention notifications about changes on your work.

CustomContextualRelevancy: The retrieval context mentions automatic watch settings and user profile notification settings but lacks details on major modifications triggering automatic page watching and deleting filters in 'Custom Filters'.

Correctness: The actual output describes how to manage notifications but lacks specific details about automatic page watching and deleting notification filters, which are present in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.9060536385671756

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Correctness: 0.9077504028316248

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and mostly addresses the question about configuring OpenID Connect in XWiki. However, there is a minor irrelevant mention of LDAP Authenticator, which slightly detracts from the focus on OpenID Connect configuration.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node, which mentions '#oidc.enableUser=true' as the necessary configuration to enable user accounts on first login, is ranked first. All irrelevant nodes, which discuss topics like domain-based authentication, session cookies, password resetting, and other unrelated authentication mechanisms, are ranked lower, ensuring precise and accurate retrieval.

ContextualRecall: The score is 0.60 because while the retrieval context's 1st node supports the instructions regarding the 'oidc.enableUser=true' property in the xwiki.properties file, it lacks explicit details about saving the file, restarting the XWiki instance, and ensuring new user accounts are automatically enabled on first login via OpenID Connect.

CustomContextualRelevancy: The retrieval context contains the exact information from the expected answer, specifically mentioning the property 'oidc.enableUser=true' in the 'xwiki.properties' file, which matches the expected output.

Correctness: The actual output accurately reflects the expected instructions for setting 'oidc.enableUser=true' and restarting XWiki. However, it does not mention uncommenting the property, which is noted in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.9145316594031131

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Correctness: 0.8871899564186785

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the first node in the retrieval context is directly related to customizing user profile fields, mentioning 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user' and providing an example of mapping. All irrelevant nodes, such as the second node discussing 'access tokens and scopes' and the third node focused on 'group synchronization and custom claims', are ranked lower, ensuring high precision. Great job maintaining perfect precision!

ContextualRecall: The score is 0.60 because while the main concept of mapping user profile fields from the OpenID Connect provider is supported by nodes in the retrieval context, specific instructions such as replacing placeholders and restarting XWiki are not explicitly found in the context.

CustomContextualRelevancy: The retrieval context contains all the information from the expected answer, including customizing user profile fields using `oidc.user.mapping` in `xwiki.properties` and the example mappings provided.

Correctness: The actual output correctly describes the customization of user profile fields through the `oidc.user.mapping` property and provides examples of mapping, following the expected output closely. However, it includes an additional note on using transformations like `.lowerCase`, which is not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8071370102121169

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.8333333333333334

ContextualRecall: 0.5

CustomContextualRelevancy: 0.8

Correctness: 0.7094887279393676

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.83 because relevant nodes discussing the `oidc.groups.mapping` and `oidc.groups.claim` properties, crucial for configuring group synchronization, are ranked higher. However, the score is not higher because nodes in retrieval context ranked third, fourth, and fifth provide general information about OpenID Connect, authentication management, and dependencies, which are not directly relevant to the group synchronization process.

ContextualRecall: The score is 0.50 because while there are supportive nodes in the retrieval context for configuring properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, key procedural steps such as locating and editing the `xwiki.properties` file, and saving and restarting XWiki are not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context includes details on configuring `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, which align with the expected output. However, it lacks explicit instructions on locating and editing the `xwiki.properties` file and restarting the XWiki instance.

Correctness: The actual output includes configuring `oidc.groups.claim`, `oidc.groups.mapping`, and `oidc.groups.separator`, aligning with expected output. However, it adds extra steps like enabling group synchronization and handling group restrictions, which are not mentioned in the expected output. The step about saving and restarting XWiki is missing.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6091088133632122

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Correctness: 0.754652880179273

Reasons:

AnswerRelevancy: The score is 0.90 because the output mostly addresses the question about setting upload restrictions, but includes links to general XWiki concepts and configurations that aren't directly related to the specific requirement of restricting attachments to images or PDFs of a certain size.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.20 because the relevant node, which 'mentions setting limits on the maximum size of an attachment and provides steps to change it', is ranked fifth. This indicates that irrelevant nodes, such as the first node discussing 'linking to external files and images', the second node providing 'a general overview of XWiki's structure and features', and the third node explaining 'how to add attachments using the XWiki interface', are ranked higher than the relevant information. These irrelevant nodes do not address configuring upload restrictions or size limits, which is why the score is not higher.

ContextualRecall: The score is 0.50 because while some nodes in the retrieval context provide relevant information about configuring xwiki.properties and setting maximum attachment size (nodes 5 and 6), they do not fully cover all the steps and details mentioned in the expected output, such as the specific line for mimetype restrictions and saving changes for subwikis.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific details on configuring xwiki.properties or setting size to 10MB as in the expected output.

Correctness: The actual output correctly details setting the maximum upload size and mentions mimetype restrictions but suggests using the Attachment Validation Application instead of directly editing the xwiki.properties file for mimetype restrictions, potentially misleading from the expected method.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.6400406063916423

Individual_scores:

AnswerRelevancy: 0.8095238095238095

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.030719828826044033

Reasons:

AnswerRelevancy: The score is 0.81 because the response mostly addresses the question about denying script rights to a space administrator, but it includes multiple irrelevant source links that do not contribute to answering the question directly.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'programming rights and their implications', but does not address the denial of script rights to a space administrator. The second node lists 'various rights and their default states', but does not mention the inability to deny script rights to a space administrator. The third node provides 'descriptions of rights', but does not address the specific issue of denying script rights to a space administrator. Subsequent nodes similarly fail to address the input query, focusing on unrelated topics such as 'comment, edit, and delete rights', 'technical implementation of access checks', and 'converting rights during page migration'.

ContextualRecall: The score is 1.00 because all sentences in the expected output are fully supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context confirms that rights implied by admin cannot be denied, matching the expected output.

Correctness: The actual output incorrectly suggests that script rights can be denied for a space administrator, contradicting the expected output that states this is not supported in XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.6409501522286888

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.7647485324197508

Reasons:

AnswerRelevancy: The score is 0.71 because the output partially addresses the input question about configuring permissions in XWiki, but includes multiple irrelevant source links that do not contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node in retrieval context, which provides information on XWiki's permission system including setting view and edit rights, is ranked sixth. The first node discusses content organization and rights inheritance in XWiki, which doesn't directly address setting view and edit permissions for teams. The second node is about the parent/child relationship and nested spaces, not directly related to configuring team-specific permissions. The third node discusses string formatting and user properties in OpenID Connect, unrelated to access control in XWiki. These irrelevant nodes should be ranked lower than the relevant node.

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and inheritance, but lacks specifics on creating groups for each team, setting wiki-wide rights, and space-level rights as described in the expected output.

Correctness: The actual output accurately describes setting global view permissions and restricting edit permissions to specific spaces, aligning with the expected output. However, it lacks mention of creating groups for each team and inheriting rights for child pages, which are key facts in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.4265580881444

Individual_scores:

AnswerRelevancy: 0.8421052631578947

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.2172432657085051

Reasons:

AnswerRelevancy: The score is 0.84 because the main issue about the notification bell not being available is addressed well, but the inclusion of source links that do not directly address the issue slightly lowers the score.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'custom notification filters and their behavior', which is not related to enabling notifications in the wiki settings. The second node lists 'various ways to receive notifications in XWiki', but again, does not mention enabling notifications in the wiki settings. Similarly, the third node describes 'various notification preferences macros', which does not provide information on enabling notifications in the wiki settings. All nodes fail to address the specific issue of enabling notifications in the wiki settings, leading to a score of 0.00.

ContextualRecall: The score is 0.50 because while the 7th node in the retrieval context mentions configuring notifications and editing the xwiki.properties file, it does not specifically mention setting 'notifications.enabled' to 'true', which is a key part of the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the 'xwiki.properties' file or setting 'notifications.enabled' to 'true'.

Correctness: The actual output provides troubleshooting steps but does not mention setting 'notifications.enabled' to 'true' in 'xwiki.properties', which is a key fact in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.32827875255765343

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.08731957416944983

Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses the issue of not receiving notifications, but it includes source links that do not directly contribute to solving the problem, which slightly reduces its relevancy.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node contains 'image links and references to various projects, which are unrelated to the topic of notifications or their settings.' Similarly, the second node 'contains image links and references to projects, which do not pertain to the topic of enabling notifications.' This pattern continues with subsequent nodes, all of which fail to address notifications, leading to a score of 0.00.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context contains no relevant information about enabling notifications or following pages and users as described in the expected output.

Correctness: The actual output does not address enabling notifications via `xwiki.properties` or following pages and users as outlined in the expected output. It provides general troubleshooting steps instead.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.3756952339036755

Individual_scores:

AnswerRelevancy: 0.8421052631578947

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.4120661402641582

Reasons:

AnswerRelevancy: The score is 0.84 because the main content effectively addresses the question about enabling notifications for one's own actions, but the inclusion of multiple source links, which do not directly contribute to answering the question, slightly detracts from the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customization of page templates, internationalization, and other features unrelated to notification settings,' which should be ranked lower. Similarly, the second node is about 'administrating wiki features like users, groups, and rights,' not relevant to notification settings. The third node mentions 'importing office documents and other features but does not relate to notification settings.' The fourth node is about 'themes, panels, and presentation settings, which do not pertain to notification settings.' Finally, the fifth node discusses 'color themes and access rights, which are not relevant to notification settings.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters mentioned in the expected output.

Correctness: Actual output provides a general method to activate notifications for own activities, but lacks specific information on disabling the filter as mentioned in the expected output. It mentions consulting documentation, yet doesn't align fully with expected factual content.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.30481990135691966

Individual_scores:

AnswerRelevancy: 0.631578947368421

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.19734046077309675

Reasons:

AnswerRelevancy: The score is 0.63 because while the output contains some relevant information about disabling notifications, it is cluttered with irrelevant URLs that do not directly address the user's question, thus lowering the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files in XWiki,' which is unrelated to managing notifications. The second node talks about 'moving attachments,' also not related to notification settings. Similarly, the third node covers 'various macros and features,' none of which pertain to notifications. This pattern continues with subsequent nodes, such as the fourth node about the 'Distribution Wizard,' which is also irrelevant. Therefore, the retrieval did not provide any relevant context for the input query.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about automatic page following or notification settings mentioned in the expected output.

Correctness: The actual output discusses managing notifications but misses key details about automatic following of pages after significant changes and lacks mention of deleting custom filters, which are present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.4097264617089868

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8333333333333334

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.22502543692058738

Reasons:

AnswerRelevancy: The score is 1.00 because the output perfectly addresses the input query without any irrelevant information. Great job!

Faithfulness: The score is 0.83 because the actual output incorrectly states that setting `xwiki.authentication.createuser=1` in the `xwiki.cfg` file ensures automatic creation of new user accounts, whereas the retrieval context specifies that it should be set to `empty` for creating users.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'domain-based authentication and session cookie configuration', which does not relate to 'automatically activating user accounts upon first login with OpenID Connect'. The second node mentions 'container authentication and various authentication methods', but fails to address the specific configuration for automatic user activation with OpenID Connect. Similarly, the

third node, despite mentioning OpenID Connect, focuses on 'authentication and synchronization', not the required configuration for automatic user activation. The fourth node provides general information about the OpenID Connect Authenticator without mentioning the configuration for automatic user activation. The fifth node discusses various OpenID Connect endpoints but omits the specific property 'oidc.enableUser' necessary for automatic user activation. The sixth node lists dependencies for the OpenID Connect extension, which is unrelated to the configuration needed. Lastly, the seventh node covers 'features and troubleshooting' for the OpenID Connect Authenticator but does not address the configuration for automatic user activation.

ContextualRecall: The score is 0.40 because while the 5th node in the retrieval context supports the mention of 'xwiki.properties' and configuration changes related to OpenID Connect, it lacks details on specific property updates, the procedure for saving and restarting XWiki, and the outcome of automatic user account activation.

CustomContextualRelevancy: The retrieval context does not mention the 'oidc.enableUser' property or its configuration in 'xwiki.properties', which is essential to the expected output.

Correctness: The actual output suggests modifying xwiki.cfg instead of xwiki.properties for automatic user activation, a discrepancy with the expected output. It also includes additional steps like checking authentication class and OpenID provider, which are not mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.5177231371656784

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.2

Correctness: 0.47300548966073686

Reasons:

AnswerRelevancy: The score is 0.83 because the output mostly addresses the input question about customizing user profile fields in XWiki, but it includes irrelevant source links and a disclaimer that do not directly contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. For example, the first node discusses 'container authentication and Active Directory integration,' which is not related to 'customizing OpenID Connect user profile fields in XWiki.' Similarly, the second node focuses on 'domain-based authentication and session cookie configuration,' which does not pertain to 'mapping additional user profile fields from OpenID Connect to XWiki.' None of the nodes provide relevant information on the specific task of customizing user profile fields using OpenID Connect in XWiki.

ContextualRecall: The score is 0.60 because while the general concept of configuring OpenID Connect in the xwiki.properties file is supported by node 7 in the retrieval context, specific examples and instructions provided in the expected output are not directly found in any nodes.

CustomContextualRelevancy: The retrieval context mentions OpenID Connect and xwiki.properties but lacks details on user mapping customization and restarting XWiki, which are key points in the expected output.

Correctness: Key facts about customizing user profile fields and using xwiki.properties are mentioned, but there is a lack of specific examples and variable mappings like in the expected output. Numerical

consistency is not applicable, and while some details like file names are consistent, specific mappings are missing. Conclusions about configuration are generally supported.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.46217461882225214

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0.20833333333333331

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.0

Correctness: 0.43138104626684576

Reasons:

AnswerRelevancy: The score is 0.80 because the output mostly addresses the process of setting up group synchronization between XWiki and an OpenID Connect provider. However, it includes some statements that are not directly relevant to the setup process, which prevents the score from being higher.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job on maintaining perfect alignment!

ContextualPrecision: The score is 0.21 because the relevant nodes in the retrieval context, specifically the sixth and eighth nodes, discuss 'Group synchronization' and the 'oidc.groups.mapping' property, which are directly relevant to the input. However, the first five nodes, which are ranked higher, are irrelevant as they discuss topics like 'authentication management and endpoints', 'container authentication', 'domain-based authentication', 'general information about OpenID Connect', and 'dependencies for the OpenID Connect extension', none of which mention group synchronization. This misranking of irrelevant nodes above relevant ones results in a lower score.

ContextualRecall: The score is 0.33 because while the 6th node in the retrieval context supports the general concept of group synchronization and configuring properties in 'xwiki.properties', it lacks specific details and examples provided in the expected output, such as opening the file with a text editor, using specific property values, and saving the file.

CustomContextualRelevancy: The retrieval context lacks specific details about configuring group synchronization properties such as 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' as outlined in the expected output.

Correctness: Actual Output omits steps regarding 'oidc.groups.claim' and separator configuration, deviating from Expected Output. It includes group mapping but lacks detail on saving 'xwiki.properties' and restarting XWiki. Adds context on custom synchronization and error checking, which aligns with expected context but lacks direct factual accuracy.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6525600122040996

Individual_scores:

AnswerRelevancy: 0.7368421052631579

Faithfulness: 1.0

ContextualPrecision: 0.26666666666666666

ContextualRecall: 0.75

CustomContextualRelevancy: 0.5

Correctness: 0.661851301294773

Reasons:

AnswerRelevancy: The score is 0.74 because the response partially addresses the input question about configuring upload settings, but the presence of multiple irrelevant source links prevents a higher score.

Faithfulness: The score is 1.00 because there are no contradictions; the actual output perfectly aligns with the retrieval context. Great job!

ContextualPrecision: The score is 0.27 because the relevant nodes in the retrieval context, such as the fifth node which includes sections on 'Size Limit' and 'Mimetype Restriction', and the sixth node which provides steps to set the maximum size of an attachment, are ranked lower than irrelevant nodes. The first node discusses 'trusted domains and PDF export customization', the second node covers 'ODT and PDF export features', and the third node lists configurations like 'WebDAV and PDF export', none of which are related to configuring attachment restrictions, yet they are ranked higher.

ContextualRecall: The score is 0.75 because while nodes in the retrieval context provide detailed steps for configuring MIME types and setting maximum upload sizes, there are missing direct mentions of specific configurations such as the exact MIME type restriction line and the overall instruction about uploading only images or PDFs with a maximum size of 10 MB.

CustomContextualRelevancy: The retrieval context mentions configuring maximum attachment size and MIME type restrictions but lacks specific details on setting MIME types in xwiki.properties and exact steps for setting size to 10 MB.

Correctness: The actual output includes the correct steps for setting maximum upload size but deviates by suggesting the Attachment Validation Application for MIME type restriction instead of editing `xwiki.properties` as in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.3236424415909411

Individual_scores:

AnswerRelevancy: 0.8125

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.1293546495456463

Reasons:

AnswerRelevancy: The score is 0.81 because the answer mostly addresses the input question about denying script rights to a space administrator, but it includes irrelevant source links that do not directly contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'form validation and JavaScript patterns,' which are unrelated to denying script rights. The second node focuses on 'templates and page creation,' not rights or permissions. The third node involves a 'custom right check class' but does not address denying script rights to an admin. The fourth node describes 'field validation and registration processes,' which are unrelated to rights management. Finally, the fifth node talks about 'executing tasks with specific user rights and the scripting API,' but it does not mention denying script rights to an admin.

ContextualRecall: The score is 0.00 because the expected output discusses rights in XWiki, which is not mentioned in any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about XWiki not supporting the denial of rights implied by the admin right.

Correctness: The actual output suggests that script rights can be denied for a space administrator, contradicting the expected output which states that rights implied by admin rights cannot be denied in XWiki.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.44601139601139606

Individual_scores:

AnswerRelevancy: 0.6538461538461539

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2222222222222222

CustomContextualRelevancy: 0.0

Correctness: 0.8

Reasons:

AnswerRelevancy: The score is 0.65 because the response partially addresses the question about configuring user permissions in XWiki, but it is lowered due to multiple irrelevant source links that do not directly relate to the task at hand.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'XWiki's database support and efficiency', which is unrelated to configuring access controls for teams. The second node is about notifications, which does not relate to access control or team permissions in XWiki. The third node covers administrating wiki features but lacks specific steps for configuring team access as required by the input. Subsequent nodes similarly fail to address the specific access control configuration needed for teams in XWiki.

ContextualRecall: The score is 0.22 because only a few sentences in the expected output, such as sentence 1 and 2, are supported by nodes in the retrieval context, specifically nodes 6 and 3. However, many other sentences, including sentence 3, 4, and the concluding statements, lack direct support from any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information about creating groups, setting rights, or managing access control in XWiki, which are all key components of the expected answer.

Correctness: The actual output aligns well with the expected output by covering global permissions, area-specific permissions, and group management. However, it lacks specific instructions for setting 'View' rights for 'XWikiAllGroup' and does not mention inheriting rights for child pages.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.7585797674180731

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 0.25
ContextualRecall: 1.0
CustomContextualRelevancy: 1.0
Correctness: 0.3570341600639933

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant to the issue of the notification bell not being available, but it includes an unnecessary link to XWiki syntax documentation, which slightly detracts from its relevance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining accuracy and alignment with the retrieval context!

ContextualPrecision: The score is 0.25 because the relevant node in the retrieval context, which mentions enabling notifications with 'notifications.enabled = true' in the xwiki.properties file, is ranked fourth. Irrelevant nodes, such as the first node discussing 'custom notification filters and their behavior', should be ranked lower as they do not address enabling notifications in the wiki.

ContextualRecall: The score is 1.00 because every part of the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context contains the exact information about enabling notifications by setting 'notifications.enabled' to 'true' in 'xwiki.properties', as found in the expected output.

Correctness: The actual output includes the key fact about setting 'notifications.enabled' to 'true', but adds extra steps not in the expected output, and sources are unrelated.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.6179858786572173
Individual_scores:
AnswerRelevancy: 0.8666666666666667
Faithfulness: 1.0
ContextualPrecision: 0.7222222222222222
ContextualRecall: 0.5714285714285714
CustomContextualRelevancy: 0.3
Correctness: 0.24759781162584357

Reasons:

AnswerRelevancy: The score is 0.87 because the response is mostly relevant to the issue of not receiving notifications, but it includes unrelated links that slightly detract from its focus.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.72 because the relevant nodes are mostly ranked higher than the irrelevant ones. The first node is relevant as it discusses 'Enable and disable the extension' and 'Enabling notifications for all users by default', which directly relates to ensuring notifications are activated. However, the second node is irrelevant since it talks about notification settings and templates without addressing enabling notifications or following pages/users, yet it is ranked higher than the third relevant node, which includes 'Watched Entities' and 'Follow a user', directly relevant to receiving notifications. Additionally, the fourth node is irrelevant as it focuses on RSS feeds and macro settings, not related to enabling notifications, but is ranked higher than the sixth relevant node discussing 'Enabling notifications for all users by default'.

ContextualRecall: The score is 0.57 because while there are nodes in the retrieval context that support the general concepts of enabling notifications and following users or pages (nodes 1 and 3), specific details such as 'xwiki.properties', receiving notifications for user-triggered events, and the 'network' tab

in the user profile are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting `notifications.enabled` to true and specific instructions on following pages or users.

Correctness: Actual Output provides comprehensive troubleshooting steps but lacks direct instructions on enabling notifications via 'xwiki.properties' as per Expected Output. It also does not mention following users for event notifications, which is a key detail missing. Some relevant context on notification settings is present but not complete.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8465926893259722

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Correctness: 0.7704652268649241

Reasons:

AnswerRelevancy: The score is 0.91 because the response is highly relevant and helpful in addressing how to enable notifications for one's own actions. However, it includes a link to XWiki syntax documentation, which is not directly related to the query, slightly affecting the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which states 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' is ranked first. This directly addresses the input question about not receiving notifications for one's own activity and how to enable them. Great job on getting it spot on!

ContextualRecall: The score is 0.50 because only the first sentence of the expected output is supported by the 1st node in the retrieval context, while the information about disabling the filter in 'Options de filtrage avancées' is not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context confirms that by default, users do not receive notifications for their own actions and mentions the ability to change this setting, aligning closely with the expected output. However, it doesn't explicitly mention the 'Options de filtrage avancées' section.

Correctness: The actual output matches most factual elements of the expected output, such as not receiving notifications for own activity by default and the ability to disable this filter. However, the actual output provides additional procedural details not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.5175052339047993

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Correctness: 0.22169807009546277

Reasons:

AnswerRelevancy: The score is 0.83 because the answer is mostly relevant, addressing the question about disabling notifications. However, it includes links and references that do not directly address the specific issue, slightly reducing its relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. Each node fails to address 'disabling automatic notifications or managing auto-watched pages'. For instance, the first node discusses 'various notification filters and settings' without mentioning how to disable automatic notifications, and the second node lists 'different notification preferences and macros' but lacks information on managing auto-watched pages. As a result, the irrelevant nodes are not ranked lower than any relevant nodes, leading to a score of 0.00.

ContextualRecall: The score is 0.75 because while nodes in the retrieval context cover aspects like default notification settings and custom filters (relating to sentences 1, 3, and 4 of the expected output), they do not explicitly mention being notified of changes to one's work, which affects sentence 2.

CustomContextualRelevancy: The retrieval context mentions auto-watched pages and notification settings but lacks details on modifying 'Suivi automatique des pages' and removing filters from 'Filtres personnalisés'.

Correctness: The actual output discusses settings for notifications but misses key details from the expected output about automatic page watching and modifying notification filters.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.46789112412820605

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 0.8571428571428571

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Correctness: 0.3057594431819349

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant to configuring OpenID Connect in XWiki, but it includes an irrelevant mention of LDAP authentication, which slightly detracts from the focus on the main topic.

Faithfulness: The score is 0.86 because the actual output suggests configuring ``xwiki.authentication.authclass=org.xwiki.contrib.oidc.auth.OIDCAuthServiceImpl`` in ``xwiki.cfg``, but this configuration is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based instance authentication and session cookie configuration,' which is unrelated to enabling user accounts upon first login using OpenID Connect. The second node mentions 'authentication management class and various OpenID Connect endpoints,' but does not provide information about enabling user accounts automatically. Similarly, the third node provides a description of the OpenID Connect authenticator and links to configuration guides, but it does not address the specific configuration needed to enable user accounts automatically.

ContextualRecall: The score is 0.50 because while there is some alignment with the retrieval context regarding the xwiki.properties file and properties like oidc.enableUser, specific instructions and effects of the modifications are not directly supported by the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions the xwiki.properties file but lacks specific details about enabling user accounts with oidc.enableUser=true.

Correctness: The actual output mentions adjusting `xwiki.properties` but omits the specific property `oidc.enableUser=true`, which is crucial for activating accounts. It also includes unrelated configuration steps, which are not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.5173066371250744

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.2

Correctness: 0.5705064894171128

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to the question about customizing user profile fields in XWiki with OpenID Connect. However, it includes a link about UI validation, which is not directly related to the query, slightly affecting the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'container authentication and various authentication methods', which does not relate to 'mapping user profile fields or customization using xwiki.properties'. The second node is about 'domain-based instance authentication and session cookie configuration', which is unrelated to the query. Subsequent nodes also fail to address the topic of 'user profile field mapping' in the context of OpenID Connect, focusing instead on unrelated aspects like 'endpoints and customization options', 'general information about the OpenID Connect Authenticator', 'dependencies for the OpenID Connect extension', and 'installation instructions'.

ContextualRecall: The score is 0.40 because while the nodes in the retrieval context mention 'OpenID Connect Authentication' and 'xwiki.properties', which support the customization and mapping of user profile fields (as seen in sentences 1 and 2 of the expected output), they lack specific examples, mapping syntax, and instructions on replacing variables, saving the file, and restarting XWiki (as detailed in sentences 3, 4, and 5 of the expected output).

CustomContextualRelevancy: The retrieval context mentions OpenID Connect and xwiki.properties but lacks details on customizing user profile fields or using oidc.user.mapping.

Correctness: The Actual Output correctly mentions modifying `xwiki.properties` for user profile mapping, but lacks specific examples for `oidc.user.mapping` and omits the restart step. It introduces additional information on using scopes and potential custom development, which is not in the Expected Output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5808312516392465
Individual_scores:
AnswerRelevancy: 0.8333333333333334
Faithfulness: 1.0
ContextualPrecision: 0.25
ContextualRecall: 0.5384615384615384
CustomContextualRelevancy: 0.5
Correctness: 0.3631926380406078
Reasons:

AnswerRelevancy: The score is 0.83 because the main content effectively addresses the process of configuring group synchronization between XWiki and an OpenID Connect provider. However, the inclusion of source links and a general suggestion to consult documentation slightly detracts from the direct relevance to the specific configuration process.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.25 because the relevant node in the retrieval context, which is the fourth node, mentions 'group synchronization and the use of the `oidc.groups.mapping` property', directly addressing the input query. However, it is ranked lower than three irrelevant nodes. The first node discusses 'authentication management class and OpenID Connect provider endpoints', the second node talks about 'container authentication and various authentication methods', and the third node is about 'domain-based authentication and session cookie configuration', none of which relate to group synchronization. This misordering results in a lower score.

ContextualRecall: The score is 0.54 because while the 4th node in the retrieval context provides detailed information on configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', it lacks references to locating and editing the 'xwiki.properties' file, replacing placeholders with actual values, and saving and restarting XWiki, which are essential steps in the expected output.

CustomContextualRelevancy: The retrieval context mentions group synchronization using 'oidc.groups.mapping' and 'xwiki.properties' but lacks details on 'oidc.groups.claim' and 'oidc.groups.separator' configurations found in the expected output.

Correctness: The actual output addresses group claims and mappings but uses a different property `oidc.userinfoclaims` instead of `oidc.groups.claim`. It lacks details on configuring `oidc.groups.separator` and saving changes, which are present in the expected output.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_028_result.json

Overall_score: 0.38016156408134877
Individual_scores:
AnswerRelevancy: 0.625
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.6559693844880929
Reasons:

AnswerRelevancy: The score is 0.62 because the actual output contains multiple irrelevant URL statements that do not address the input question about configuring file upload restrictions, which lowers the score. However, the score is not lower because there is still some relevant information provided regarding the configuration of file uploads.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query about file upload restrictions. For example, the first node discusses 'language preferences and translations in XWiki', which is unrelated to file upload settings. Similarly, the second node covers 'mail configuration and theme customization', which also does not address file upload settings.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output could be attributed to any nodes in the retrieval context, indicating a complete mismatch between the expected information and the available context.

CustomContextualRelevancy: The retrieval context lacks any information related to configuring MIME types or setting maximum attachment size, which are key components of the expected output.

Correctness: The actual output properly addresses MIME type configuration and attachment size, similar to the expected output, but omits specific steps for changing settings via the XWikiPreferences interface and does not mention repeating the process for sub-wikis.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.5508607628053123

Individual_scores:

AnswerRelevancy: 0.6470588235294118

Faithfulness: 0.8

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.15810575330246157

Reasons:

AnswerRelevancy: The score is 0.65 because the response partially addresses the question about refusing script rights to a space administrator, but it is lowered by multiple irrelevant source link statements that do not contribute to answering the question directly.

Faithfulness: The score is 0.80 because the actual output incorrectly suggests that script rights are allowed by default at the main wiki level, whereas the retrieval context clarifies that for XWiki 14.10+, script rights are not given by default to all users at the main wiki level.

ContextualPrecision: The score is 0.50 because relevant nodes in the retrieval context are ranked lower than irrelevant nodes. The first node is about 'language preferences and translations in XWiki', which is unrelated to script rights or administrative permissions, and should be ranked lower. Similarly, the third node discusses 'interacting with the XWiki RESTful API', which is also unrelated and should be ranked lower. The second node, which is relevant, discusses the 'Script' right and its allowance or denial, and should be ranked higher.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the retrieval context, showcasing a comprehensive understanding of the topic. Great job!

CustomContextualRelevancy: The retrieval context mentions that certain rights like ADMIN are not deniable, but it does not explicitly state that administrator rights cannot be refused, which is the key information in the expected output.

Correctness: Actual output provides steps to refuse script rights, whereas expected output states it's not supported in XWiki. No factual consistency and contains extra information.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.40779402876098875

Individual_scores:

AnswerRelevancy: 0.7058823529411765

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.7408818196247561

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about configuring access rights in XWiki, but it includes multiple irrelevant source links that do not directly answer the query.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job maintaining perfect alignment!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question. For example, the first node discusses 'the structure of an XWiki page' without addressing access controls, and the second node covers 'modifying resources using XML and HTTP methods', which is unrelated to configuring access rights in XWiki.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks details on creating groups, setting wiki-level and space-level rights, and inheriting rights for child pages, which are crucial steps in the expected output.

Correctness: The actual output aligns with the expected output regarding access rights configuration, ensuring read access for all and edit rights restricted to team members. However, it lacks specific instructions about creating groups for each team and inheriting rights for child pages, which are mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7537122941725168

Individual_scores:

AnswerRelevancy: 0.6875

Faithfulness: 0.8333333333333334

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.5014404317017673

Reasons:

AnswerRelevancy: The score is 0.69 because the output includes relevant information about authenticating users with access tokens, but it is lowered due to repeated irrelevant statements providing source links that do not directly address the question.

Faithfulness: The score is 0.83 because the actual output incorrectly suggests setting `xwiki.authentication.authclass` to `org.xwiki.contrib.ilm.authentication.JWTTokenAuth` for JWT Token Authentication, whereas the retrieval context specifies it should be set to `com.xpn.xwiki.user.impl.xwiki.AppServerTrustedAuthServiceImpl` for container authentication.

ContextualPrecision: The score is 0.50 because relevant nodes in the retrieval context are not consistently ranked higher than irrelevant nodes. For example, the first node is about 'LDAP

configuration' and does not mention token-based authentication, which should be ranked lower. Similarly, the third node is about 'LDAP authentication' and is not related to token-based authentication, yet it appears before more relevant nodes. On the other hand, relevant nodes like the second node, which discusses 'configuration of authorized applications' for token-based authentication, and the fourth node, which details 'JWT token generation', are appropriately included but not consistently prioritized over irrelevant nodes.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the relevant nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context contains detailed information about token-based authentication, JWT claims, and configuration steps, matching the expected answer.

Correctness: The actual output mentions JWT Token Authentication and setting `xwiki.authentication.authclass`, aligning with the expected output. However, it introduces OpenID Connect which is not in the expected output, and lacks details on configuring authorized applications and claim requirements.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.31899531887666716

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.16397191326000277

Reasons:

AnswerRelevancy: The score is 0.75 because the response addresses the question about authenticating users with access tokens but includes multiple irrelevant source links, which do not contribute to answering the question directly.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job on maintaining perfect alignment!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'model validation strategies in XWiki' without mentioning token-based authentication. The second node talks about 'attachment validation and configuration in XWiki,' which is unrelated to user authentication with access tokens. The third node details 'the admin UI for model validation,' not relevant to authenticating users with access tokens. The fourth node is about 'setting up fields and validation for user registration in XWiki,' which does not pertain to token-based user authentication. Lastly, the fifth node describes 'steps for installing a UI in XWiki,' which is unrelated to the process of authenticating users with access tokens.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are key components of the expected output.

Correctness: The actual output generally discusses token-based authentication but lacks specific details such as JWT specifics, Ed25519 signing, and configuration steps present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.3314417551954443

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.20293624545838002

Reasons:

AnswerRelevancy: The score is 0.79 because the response largely addresses the question about authenticating users with access tokens, but it includes repeated irrelevant source links that do not directly answer the query, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node is about 'upgrading XWiki' and does not mention authentication or access tokens. The second node discusses 'features related to users and rights' but not authentication with access tokens. The third node mentions 'user authentication' but lacks information on using access tokens. The fourth node is about 'rights management' and is unrelated to authentication with access tokens. Lastly, the fifth node lists 'different rights and their properties', which is also irrelevant to the use of access tokens for authentication.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, as the context lacks any mention of key concepts such as 'authentifier les utilisateurs avec des jetons d'accès', 'applications externes', or 'jetons Web JSON signés (JWT)'.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which is the focus of the expected answer.

Correctness: The actual output does not address the specific requirement of using the JWTTOKENAuth extension for authentication as outlined in the expected output. It mentions potential support through extensions but lacks details on activating the authenticator or configuring allowed applications.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_GPT4o-mini

File: qa_001_result.json

Overall_score: 0.45158911416588193

Individual_scores:

AnswerRelevancy: 0.8518518518518519

Faithfulness: 1.0

ContextualPrecision: 0.30952380952380953

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.04815902361963014

Reasons:

AnswerRelevancy: The score is 0.85 because the response is mostly relevant to the issue of the missing 'Bell' icon in XWiki, but it includes repeated irrelevant statements about source links that do not

directly address the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.31 because the relevant nodes in the retrieval context are ranked lower than several irrelevant nodes. For instance, the first node discusses 'installation instructions and prerequisites', which is not relevant to fixing the bell issue, and the second node is about 'access and error messages related to configurable applications', which also does not pertain to the problem. The third node, which is relevant as it mentions the 'Notifications Application' and the bell icon, is ranked after these irrelevant nodes. Similarly, the seventh node, which is relevant because it discusses 'configuring notification settings in the administration', is also ranked lower than irrelevant nodes like the fourth node, which only describes 'the content of the Alert menu and notifications' without addressing the bell issue.

ContextualRecall: The score is 0.50 because while the 7th node in the retrieval context mentions 'notifications' and 'xwiki.properties', there is no specific mention of enabling notifications via 'xwiki.properties', which is crucial for fully supporting the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the 'xwiki.properties' file, which is the key information in the expected output.

Correctness: The actual output does not mention enabling notifications via 'notifications.enabled' in 'xwiki.properties', and introduces unrelated troubleshooting steps.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6841433173428224

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.6

Correctness: 0.3381932373902682

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and addresses the question without any irrelevant statements. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.67 because the first node in the retrieval context is relevant, explaining how to follow a user and receive notifications, which is crucial for ensuring notifications are set up correctly. However, the second node discusses 'filters and hidden pages,' which is not directly related to enabling notifications, and should be ranked lower. Additionally, the third node focuses on 'customizing the notification email template,' which is not directly relevant to the issue of not receiving notifications, and should also be ranked lower than relevant nodes.

ContextualRecall: The score is 0.50 because while the retrieval context supports details about watching pages, following users, and using the 'network' tab (as seen in node 1 of the retrieval context), it lacks information about enabling notifications in 'xwiki.properties' and subscribing to pages or users for notifications.

CustomContextualRelevancy: The retrieval context covers enabling notifications and following users, but lacks details on setting 'notifications.enabled' in 'xwiki.properties' and subscribing to pages.

Correctness: The actual output mentions enabling notifications and watching users or pages, but lacks specific instructions on editing 'xwiki.properties' and does not address using the 'alert' menu or

checking the 'network' tab.
Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_003_result.json

Overall_score: 0.9496539698274892
Individual_scores:
AnswerRelevancy: 0.7692307692307693
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 1.0
CustomContextualRelevancy: 1.0
Correctness: 0.9286930497341664
Reasons:

AnswerRelevancy: The score is 0.77 because the response mostly addresses the question about enabling notifications for one's own actions in XWiki. However, the inclusion of source links that do not directly answer the question slightly detracts from the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant ones. The first two nodes directly address the input by explaining how to enable notifications for one's own actions, while the irrelevant nodes discuss unrelated topics such as following a user, administrator settings, RSS feeds, and notification methods, which are not pertinent to the input question. Great job on the perfect ranking!

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context fully contains the expected answer, including the default behavior of not receiving notifications about one's own activity unless explicitly targeted, and the ability to disable this filter in the 'Advanced filtering options' section of the user profile.

Correctness: The actual output accurately reflects the expected output by explaining that users do not receive notifications for their own actions by default and provides instructions on how to disable the filter in the 'Advanced filtering options' of notification settings, matching the expected output details.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_004_result.json

Overall_score: 0.5707448917994089
Individual_scores:
AnswerRelevancy: 0.8947368421052632
Faithfulness: 1.0
ContextualPrecision: 0.26785714285714285
ContextualRecall: 0.75
CustomContextualRelevancy: 0.3
Correctness: 0.21187536583404754
Reasons:

AnswerRelevancy: The score is 0.89 because the response mostly addresses how to disable notifications in XWiki, but includes a source link and a standalone 'Filters:' statement that do not directly contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.27 because the relevant nodes in the retrieval context are ranked lower than several irrelevant nodes. For instance, the first node discusses 'following a user and default watch settings,' which does not address disabling notifications, while the fourth node, which is relevant, mentions 'administrators can change users' notification settings.' Additionally, the second node focuses on 'details of changes and macros for settings,' which is not pertinent, whereas the seventh node, which is relevant, discusses 'filtering notifications, including advanced filtering options.' These relevant nodes should be ranked higher to improve the score.

ContextualRecall: The score is 0.75 because most of the expected output is supported by node 5 in the retrieval context, which covers automatic page watching, changing settings, and custom filters. However, the retrieval context does not mention notifications about changes to your work specifically, leading to a partial match.

CustomContextualRelevancy: The retrieval context mentions default watch settings and notification settings in user profiles, but lacks details about automatic page watching for major modifications and deleting notification filters.

Correctness: The actual output provides steps for managing notifications but lacks key facts about automatic page watching and deleting custom filters from the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8990661087973559

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Correctness: 0.8777299861174692

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant to configuring OpenID Connect in XWiki, but it includes an irrelevant mention of LDAP authentication, which slightly detracts from the overall focus.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which discusses the property '# oidc.enableUser=true' necessary for enabling user accounts on first login, is ranked first. This ensures that the most pertinent information is prioritized over irrelevant nodes discussing topics like domain-based authentication, password recovery, and general OpenID Connect configurations.

ContextualRecall: The score is 0.60 because while the 1st node in the retrieval context supports the instructions to adjust the 'xwiki.properties' file and set the 'oidc.enableUser' property to 'true', it lacks information about saving the file, restarting the XWiki instance, and ensuring new user accounts are automatically enabled on first login via OpenID Connect.

CustomContextualRelevancy: The retrieval context contains the exact information from the expected output, specifically mentioning the property 'oidc.enableUser=true' in the 'xwiki.properties' file, which matches the expected answer.

Correctness: The actual output correctly describes adjusting the 'oidc.enableUser' property, saving the file, and restarting XWiki. It omits the specific commenting/uncommenting detail present in the expected

output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.8490539647125716

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 1.0

ContextualPrecision: 0.625

ContextualRecall: 0.8

CustomContextualRelevancy: 1.0

Correctness: 0.7869708470989581

Reasons:

AnswerRelevancy: The score is 0.88 because the response is mostly relevant to the question about customizing user profile fields in XWiki using OpenID Connect. However, it includes citations and suggestions to consult documentation, which do not directly address the specific customization query, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.62 because the first node in the retrieval context is relevant and directly addresses customizing user profile fields with an example of mapping, which is highly pertinent to the input. However, several irrelevant nodes, such as the second node discussing 'access tokens and scopes', the third node about 'group claims and synchronization', and others, are ranked higher than the last relevant node, which lists available variables for mapping. These irrelevant nodes should be ranked lower to improve the precision score.

ContextualRecall: The score is 0.80 because the retrieval context nodes support most of the expected output, particularly the customization and mapping of user profile fields in the `xwiki.properties` file (sentence 1-4). However, it lacks information on saving the file and restarting XWiki to apply changes (sentence 5).

CustomContextualRelevancy: The retrieval context contains the exact information from the expected output, including the ability to map user profile fields using the `oidc.user.mapping` property in the `xwiki.properties` file with examples provided.

Correctness: The actual output accurately describes customizing user profile fields using `oidc.user.mapping` and provides example mappings. However, it includes additional information about using custom claims and consulting documentation, which is not present in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.7763624843120526

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.8333333333333334

ContextualRecall: 0.5

CustomContextualRelevancy: 0.8

Correctness: 0.5248415725389818

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.83 because the relevant nodes in the retrieval context are mostly ranked higher than the irrelevant ones. The first and second nodes are relevant as they discuss 'oidc.groups.mapping' and group synchronization, which are crucial for configuring group synchronization between XWiki and an OpenID Connect provider. However, the third node, which provides general information about OpenID Connect without addressing group synchronization, should be ranked lower. Similarly, the fourth node focuses on authentication management and endpoints, and the fifth node lists dependencies, both of which are not directly related to the group synchronization process. The sixth node is relevant as it describes the 'oidc.groups.claim' property, essential for the setup, and is appropriately ranked.

ContextualRecall: The score is 0.50 because while several specific configuration steps (sentences 3, 5, and 7) are supported by nodes in the retrieval context, many general instructions and replacement examples (sentences 1, 2, 4, 6, 8, and 9) are not directly supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context includes details on configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', which align with the expected output. However, it lacks explicit instructions on locating and editing the 'xwiki.properties' file and restarting the XWiki instance.

Correctness: The actual output mentions configuring 'oidc.groups.claim' and 'oidc.groups.mapping', aligning with expected output. However, it lacks details on 'oidc.groups.separator' and saving/restarting XWiki. Additional steps like enabling group synchronization and testing are present but not in expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6273544245715755

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.5833333333333334

CustomContextualRelevancy: 0.3

Correctness: 0.6807932140961195

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.20 because the relevant node in the retrieval context, which 'mentions setting limits on the maximum size of an attachment and provides steps to change it', is ranked fifth. However, several irrelevant nodes are ranked higher, such as the first node, which 'discusses linking to external files and images, which is unrelated to configuring upload restrictions for attachments', and the second node, which 'provides general information about XWiki pages and their components, not about configuring attachment upload settings'.

ContextualRecall: The score is 0.58 because while several steps related to configuring maximum attachment size (Step 2) are well-supported by node 5 in the retrieval context, key instructions regarding mimetype configuration (Step 1) and ensuring only specific file types and sizes are allowed lack direct support from any node in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and configuring mimetype restrictions, but lacks specifics on editing xwiki.properties for mimetypes and setting 10MB limit.

Correctness: The actual output correctly identifies the need to set a 10MB upload limit and specify allowed MIME types but suggests using the Attachment Validation Application instead of directly editing `xwiki.properties` for MIME types. It also includes additional steps like security settings and testing, which are not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.7189866824745721

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0.7

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Correctness: 0.01392009484743239

Reasons:

AnswerRelevancy: The score is 0.80 because the answer mostly addresses the question about denying script rights to a space administrator, but it includes multiple irrelevant source links that do not directly pertain to the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.70 because the first node in the retrieval context is relevant, as it mentions that 'programming rights can only be granted from the wiki preferences page,' implying that certain rights, including admin rights, cannot be denied once granted. However, the second node is irrelevant, as it primarily lists various rights and their default states but does not specifically address the denial of script rights to a space administrator. The third node is also irrelevant, providing a tabular view of rights but not mentioning the inability to deny script rights to a space administrator. These irrelevant nodes are ranked higher than the fifth node, which is relevant and explains that 'the script right gives a lot of power to users' and that administrators have to manually allow it, suggesting that once allowed, it cannot be denied.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context mentions that admin rights imply other rights and are not deniable, aligning with the expected output that rights implied by admin cannot be denied. However, it does not explicitly state that denying admin rights is unsupported in XWiki.

Correctness: The actual output provides steps to deny script rights to a space administrator in XWiki, contradicting the expected output which states it is not supported and rights implied by admin right cannot be denied.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.6467811474647401

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 0.8888888888888888

ContextualPrecision: 0.20833333333333331

ContextualRecall: 0.875

CustomContextualRelevancy: 0.3

Correctness: 0.8084646625662186

Reasons:

AnswerRelevancy: The score is 0.80 because the response effectively addresses the main question about configuring access rights in XWiki, but the inclusion of multiple source links, which are not directly relevant to the configuration process, slightly detracts from the overall relevancy.

Faithfulness: The score is 0.89 because the actual output incorrectly states that space-level permissions override global permissions, whereas the retrieval context clarifies that page-level permissions can override wiki-wide permissions.

ContextualPrecision: The score is 0.21 because the relevant nodes in the retrieval context are ranked lower than several irrelevant nodes. The first node discusses 'the organization of content in XWiki and the concept of nested pages,' which does not provide guidance on setting view and edit rights for teams. The second node explains 'the parent/child relationship and nested spaces,' but does not address configuring access rights for teams. The third node is about 'string formatting and user identifiers,' which is unrelated to the input question. The sixth node, which is relevant, provides information on XWiki's permission system, including setting view and edit rights. The eighth node, also relevant, outlines basic rules for setting rights in XWiki, including creating groups and setting page-level rights. These relevant nodes should be ranked higher to improve the score.

ContextualRecall: The score is 0.88 because most of the expected output sentences are well supported by the nodes in the retrieval context, particularly node 6, which covers key aspects like creating groups, setting wiki-wide and space-level rights, and ensuring view and edit permissions. However, the retrieval context lacks information on adjusting rights for additional teams or spaces, which slightly impacts the score.

CustomContextualRelevancy: The retrieval context mentions rights inheritance and setting rights at different levels, but lacks specific steps like creating groups for each team or setting 'Edit' rights for team spaces.

Correctness: The actual output accurately describes setting global view rights and specific edit rights, ensuring teams can view all spaces but only edit their own. However, it lacks specific mention of creating groups for each team and inheriting rights for child pages as outlined in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.43680401910387645

Individual_scores:

AnswerRelevancy: 0.9565217391304348

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.16430237549282398

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant to the input, addressing the issue with the notification bell. However, it includes a link to XWiki syntax documentation, which is not directly related to the problem described, slightly affecting the score.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job on maintaining perfect alignment!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'custom notification filters and their behavior but does not address enabling notifications in the wiki settings,' which is not pertinent to the input query about the availability of the notification bell. Similarly, the second node lists 'various ways to receive notifications but does not mention enabling notifications through the xwiki.properties file,' which again does not relate to the input issue. As a result, there are no relevant nodes ranked higher than irrelevant ones.

ContextualRecall: The score is 0.50 because while the 6th node in the retrieval context suggests that notifications can be configured, there is no specific mention in any node about setting 'notifications.enabled' in 'xwiki.properties' to 'true', which is a key detail in the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the 'xwiki.properties' file or setting 'notifications.enabled' to 'true', which is the key information in the expected output.

Correctness: The actual output includes information about troubleshooting notification issues and does not directly address enabling notifications via 'xwiki.properties'. It mentions user permissions and extension compatibility, which are not in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3274774768533072

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.07012801901457995

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant to the issue of not receiving notifications, but it includes links to unrelated topics like XWiki skins and a security module extension, which slightly detract from its focus.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first four nodes contain 'image links and references which are not related to notifications or their settings', and the fifth node 'describes a security module and its ratings, which is unrelated to notification settings or issues'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context contains image links and unrelated text, lacking any information about enabling notifications or following pages/users as described in the expected output.

Correctness: The actual output lacks detailed instructions for enabling notifications in 'xwiki.properties' and following pages, while focusing on troubleshooting steps instead.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.3561232274243851

Individual_scores:

AnswerRelevancy: 0.8421052631578947

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2946341013884163

Reasons:

AnswerRelevancy: The score is 0.84 because the main content effectively addresses the question about enabling notifications for one's own actions, but the inclusion of multiple source links that do not directly contribute to the solution slightly detracts from the overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are unrelated to the input about enabling notifications for user actions. The first node discusses 'customization of page templates, internationalization, accessibility, and performance', the second node is about 'administrating wiki features like users, groups, and rights', and subsequent nodes cover topics such as 'importing office documents, icon themes, sharing pages, and syntax highlighting', 'themes, panels, and presentation customization', and 'color themes, access rights, and permissions'. None of these are relevant to notification settings, leading to a low score.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters, which are mentioned in the expected output.

Correctness: The actual output provides steps to enable notifications, while the expected output states they are off by default and can be disabled via 'Erweiterte Filteroptionen'. There's limited alignment on the core fact about default notification settings and the user profile setting.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.2873527554255937

Individual_scores:

AnswerRelevancy: 0.5625

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.161616532553562

Reasons:

AnswerRelevancy: The score is 0.56 because the output contains several irrelevant links that do not address the question about disabling notifications, which lowers the score. However, some relevant information may still be present, contributing to the score being above zero.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files', which does not relate to 'notification settings or disabling notifications'. The second node talks about 'moving and deleting attachments', which is unrelated to 'managing notifications'. Similarly, the third node lists 'various macros and their functions', none of which pertain to 'notification settings'. This pattern continues with subsequent nodes, such as the fourth node about the 'Distribution Wizard and its features', which do not include 'notification settings', and the fifth node providing 'examples of linking in XWiki', which is not relevant to 'notification settings'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about automatic page following or notification settings as described in the expected output.

Correctness: The actual output discusses disabling notifications, while the expected output is about automatic page following and notification settings. There is a lack of alignment in content focus and factual elements, with no mention of custom filters or automatic tracking as in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.40767889693020903

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.25

CustomContextualRelevancy: 0.0

Correctness: 0.1960733815812542

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and directly addresses the configuration issue with OpenID Connect in XWiki without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job maintaining consistency!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based authentication and session cookie configuration', which is unrelated to 'automatically activating user accounts with OpenID Connect in XWiki'. The second node provides information on 'container authentication and various authentication methods', but does not mention the specific configuration needed. Subsequent nodes continue to lack relevance, such as the third node which describes the 'OpenID Connect Authenticator extension and its capabilities', but does not provide the specific configuration details needed for automatic user account activation.

ContextualRecall: The score is 0.25 because while the 5th node in the retrieval context mentions the `xwiki.properties` file and properties related to OpenID Connect, none of the nodes provide specific information about automatically activating user accounts on first login or updating the ``oidc.enableUser`` property.

CustomContextualRelevancy: The retrieval context does not mention the ``oidc.enableUser`` property or the need to update ``xwiki.properties`` to activate user accounts automatically.

Correctness: The actual output incorrectly refers to the ``xwiki.cfg`` file and the ``xwiki.authentication.createuser`` setting, while the expected output specifies modifying ``xwiki.properties`` and setting ``oidc.enableUser=true`` for automatic user activation.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.650978378899855

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.6503147178435742

Reasons:

AnswerRelevancy: The score is 0.89 because the response mostly addresses the question about customizing user profile fields in XWiki using OpenID Connect, but it includes irrelevant source links that do not directly contribute to the solution.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node in the retrieval context, which is the sixth node, mentions the `xwiki.properties` file and includes examples of properties related to OpenID Connect, aligning with the expected output about mapping user profile fields using `oidc.user.mapping`. However, it is ranked lower than irrelevant nodes, such as the first node, which discusses various authentication methods but does not mention customizing user profile fields or mapping properties in `xwiki.properties`. Additionally, the second node is about domain-based authentication and session cookies, which is unrelated to the input question. These irrelevant nodes are ranked higher, reducing the overall precision score.

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the information in the nodes in retrieval context. Great job!

CustomContextualRelevancy: The retrieval context mentions the xwiki.properties file and OpenID Connect, but lacks details on customizing user profile fields or using oidc.user.mapping as described in the expected output.

Correctness: Actual output covers key steps for configuring OpenID in XWiki and mentions the use of `xwiki.properties`, but lacks specific mapping examples and details about restarting XWiki as in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.4589173950101399

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0.125

ContextualRecall: 0.26666666666666666

CustomContextualRelevancy: 0.2

Correctness: 0.2729488145052834

Reasons:

AnswerRelevancy: The score is 0.89 because the answer is mostly relevant and provides useful information, but it includes statements that offer sources and links without directly addressing the specific process of setting up group synchronization, which slightly detracts from its overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on ensuring the actual output aligns perfectly with the retrieval context!

ContextualPrecision: The score is 0.12 because the relevant node in the retrieval context, which describes the group synchronization process and mentions the use of the 'oidc.groups.mapping' property, is ranked eighth. However, irrelevant nodes, such as the first node focusing on 'authentication management and endpoints' without mentioning group synchronization, and the second node discussing 'container authentication and LDAP', are ranked higher, which lowers the score.

ContextualRecall: The score is 0.27 because while some nodes in the retrieval context mention relevant properties like 'xwiki.properties', 'oidc.groups.mapping', and 'oidc.groups.separator', many specific details and examples from the expected output are missing, such as instructions for opening the file, specific configuration examples, and steps for saving and restarting XWiki.

CustomContextualRelevancy: The retrieval context mentions group synchronization and configuration in xwiki.properties but lacks specific details on configuring oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping as outlined in the expected output.

Correctness: Actual output lacks specific instructions for configuring 'oidc.groups.claim' and 'oidc.groups.separator'. It mentions 'oidc.userinfoclaims' and 'oidc.groups.mapping' instead, which is inconsistent with expected output. The sequence and context around file editing and restarting XWiki are missing.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6158609252782811

Individual_scores:

AnswerRelevancy: 0.7058823529411765

Faithfulness: 1.0

ContextualPrecision: 0.26666666666666666

ContextualRecall: 0.7

CustomContextualRelevancy: 0.3

Correctness: 0.7226165320618432

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the input question about configuring upload restrictions, but it includes multiple irrelevant source links that do not directly contribute to the solution.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.27 because the relevant nodes in the retrieval context, such as the fifth node mentioning 'Size Limit' and 'Mimetype Restriction', and the sixth node providing steps to set the maximum size of an attachment, are ranked lower than irrelevant nodes. The first node discusses trusted domains and PDF export customization, which are not relevant to configuring attachment restrictions, and should be ranked lower.

ContextualRecall: The score is 0.70 because while nodes in the retrieval context provide relevant information about 'Size Limit', 'Mimetype Restriction', and steps to change 'Maximum Upload Size', there is no specific mention of configuring MIME types or blocking other MIME types, which affects the completeness of the expected output.

CustomContextualRelevancy: The retrieval context mentions configuring attachment size and MIME type restrictions but lacks specific details on setting MIME types in xwiki.properties and the exact steps for setting size to 10MB.

Correctness: Actual output covers setting maximum upload size and MIME type restrictions but mentions 'Attachment Validation Application' not in expected output, and lacks precise steps for MIME

configuration in `xwiki.properties`.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.31525845979743494

Individual_scores:

AnswerRelevancy: 0.8421052631578947

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.04944549562671495

Reasons:

AnswerRelevancy: The score is 0.84 because the main content effectively addresses the question about denying script rights to a space administrator, but the inclusion of source links, which do not directly contribute to the answer, slightly reduces the relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job maintaining accuracy!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'form validation and JavaScript validation patterns,' which are unrelated to denying script rights to a space administrator in XWiki. The second node is about 'template providers and page creation in XWiki,' which does not relate to script rights. The third node describes a 'custom right check class and configuration,' but does not address denying script rights. The fourth node focuses on 'user registration fields and validation,' which is not relevant to script rights. Finally, the fifth node discusses 'access rights and changing the current author,' but does not specifically address denying script rights to a space administrator.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about XWiki not supporting the denial of rights implied by admin rights, which is the key point in the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information about XWiki not supporting the denial of rights implied by admin rights.

Correctness: The actual output provides steps to revoke script rights, which contradicts the expected output stating it is not supported in XWiki. It misses the key fact that rights implied by admin rights cannot be denied.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.45263056174041916

Individual_scores:

AnswerRelevancy: 0.625

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2857142857142857

CustomContextualRelevancy: 0.0

Correctness: 0.8050690847282294

Reasons:

AnswerRelevancy: The score is 0.62 because while the response attempts to address configuring team access and editing permissions in XWiki, it includes multiple irrelevant source links that do not directly relate to the task, reducing its overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'XWiki's database efficiency and programming features', which are unrelated to configuring access controls for teams. The second node is about 'notifications and application details', not about setting access permissions for teams. The third node describes an 'administration application', but does not provide information on configuring team-specific access rights. Subsequent nodes continue this pattern, discussing topics like 'model validation', 'nested pages and subwikis', and 'LDAP group configurations', none of which address the specific access control configuration needed in the input.

ContextualRecall: The score is 0.29 because while nodes in the retrieval context mention managing users, groups, and rights, they do not cover specific actions like setting area-level rights, inheriting rights for subpages, or adjusting rights for additional teams, which are detailed in the expected output sentences 2, 3, 4, and the concluding statements.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting wiki-wide or area-level rights, or inheriting rights for subpages as described in the expected answer.

Correctness: The actual output closely aligns with the expected output by outlining steps to create groups, adjust rights, and ensure visibility and edit permissions for team areas. However, it does not explicitly mention inheriting rights for sub-pages, which is a part of the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.68662031071963

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Correctness: 0.21257900717492278

Reasons:

AnswerRelevancy: The score is 0.86 because the response is mostly relevant to the issue of the notification bell not being available, but it includes links that do not directly address this specific problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.25 because the relevant node in the retrieval context, which 'provides information on enabling notifications by setting `notifications.enabled` in `xwiki.properties`, which directly addresses the expected output', is ranked fourth. This indicates that several irrelevant nodes, such as the first node that 'discusses custom notification filters and their behavior, which does not relate to enabling notifications in the wiki settings', are ranked higher than the relevant one.

ContextualRecall: The score is 1.00 because the expected output perfectly matches the relevant information in the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties with 'notifications.enabled = true', which aligns with the expected answer about activating notifications

by setting 'notifications.enabled' to 'true'. However, it does not explicitly mention setting it in 'xwiki.properties' for notifications, which slightly reduces the score.
Correctness: The actual output does not address enabling notifications via 'xwiki.properties' and includes additional steps not mentioned in the expected output.
Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_022_result.json

Overall_score: 0.6673605382554418
Individual_scores:
AnswerRelevancy: 0.8823529411764706
Faithfulness: 1.0
ContextualPrecision: 0.7222222222222222
ContextualRecall: 0.7142857142857143
CustomContextualRelevancy: 0.5
Correctness: 0.18530235184824306
Reasons:
AnswerRelevancy: The score is 0.88 because the response mostly addresses the issue of not receiving notifications, but includes irrelevant source links that do not directly relate to the problem.
Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!
ContextualPrecision: The score is 0.72 because relevant nodes in the retrieval context, such as the first node mentioning 'Enable and disable the extension' and 'Enabling notifications for all users by default', are ranked higher. However, irrelevant nodes like the second node, which focuses on 'email frequency and template customization', and the fourth node, discussing 'RSS feeds and notification grouping', are not directly related to enabling notifications and should be ranked lower.
ContextualRecall: The score is 0.71 because while nodes in the retrieval context provide support for activating notifications and following pages or users (as seen in sentences 1, 3, 4, and 5 of the expected output), they lack specific references to the 'notifications.enabled' property and the 'réseau' tab (related to sentences 2 and 6).
CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users or pages, but lacks details on setting 'notifications.enabled' to 'true' and specific steps for following users or pages.
Correctness: The actual output provides general troubleshooting steps for notifications but lacks specific instructions on enabling notifications in 'xwiki.properties' or following pages/users as mentioned in the expected output.
Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8208392822628409
Individual_scores:
AnswerRelevancy: 0.8
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 0.5
CustomContextualRelevancy: 0.9
Correctness: 0.7250356935770447

Reasons:

AnswerRelevancy: The score is 0.80 because the response mostly addresses the question about enabling notifications for one's own actions, but includes irrelevant source links that do not directly relate to the specific issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which states, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' is ranked first and directly addresses the input question. All irrelevant nodes, such as the second node discussing 'getting a notification RSS feed and settings,' are ranked lower, ensuring high contextual precision. Great job on ranking the relevant information at the top!

ContextualRecall: The score is 0.50 because the first sentence is supported by the 1st node in the retrieval context, but the second sentence lacks support as none of the nodes mention the option to disable the filter in the specified settings section.

CustomContextualRelevancy: The retrieval context mentions that by default, users do not receive notifications for their own actions and can change this setting, aligning closely with the expected output. However, it does not explicitly mention 'Options de filtrage avancées' for disabling the filter.

Correctness: The actual output correctly states that users do not receive notifications for their own activity by default and describes how to change this setting. However, it omits the term 'Options de filtrage avancées' from the expected output and includes additional procedural steps and source links not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.585614193297565

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 0.8333333333333334

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.14701849311872278

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses how to disable notifications, but includes irrelevant source links that do not directly provide the necessary steps.

Faithfulness: The score is 0.83 because the actual output incorrectly claims that users do not receive notifications for their own activities by default, whereas the retrieval context specifies that users do not receive notifications about their own activity unless the event targets the user explicitly.

ContextualPrecision: The score is 0.50 because the second node in the retrieval context, which is relevant, is ranked lower than the first node. The first node discusses 'technical aspects of notification filters and hidden pages,' which are not directly related to disabling notifications for pages of no interest. This irrelevant node should be ranked lower than the second node, which mentions 'Notifications Auto Watch Preferences' and 'Notifications Custom Filters Preferences,' aligning with the input's need to modify settings related to automatically watched pages and custom filters.

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context, indicating a flawless match. Great job!

CustomContextualRelevancy: The retrieval context mentions 'auto watched pages' and 'custom notification filters', but lacks details on automatic addition to watchlist and modifying notification settings as described in the expected output.

Correctness: The actual output describes steps for disabling notifications, which is not mentioned in the expected output. While both discuss notification settings, the actual output lacks the specific details on 'Suivi automatique des pages' and 'Filtres personnalisés' as outlined in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.40394111357652634

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 0.4

ContextualPrecision: 0.5

ContextualRecall: 0.2

CustomContextualRelevancy: 0.2

Correctness: 0.19507525288772903

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to configuring OpenID Connect in XWiki, with only a minor irrelevant mention of LDAP Authenticator, which does not significantly detract from the overall helpfulness of the answer.

Faithfulness: The score is 0.40 because the actual output incorrectly associates the setting `xwiki.authentication.createuser=empty` with OpenID Connect, while the retrieval context clearly states it is for Container Authentication. This misalignment occurs multiple times, indicating a significant deviation from the retrieval context.

ContextualPrecision: The score is 0.50 because the second node in the retrieval context is relevant as it 'mentions the xwiki.properties file, which is directly related to the configuration needed to enable user accounts automatically with OpenID Connect.' However, the first node is irrelevant since it 'discusses domain-based authentication and session cookies, which are not relevant to enabling user accounts automatically on first login with OpenID Connect.' This misplacement of an irrelevant node before a relevant one affects the score.

ContextualRecall: The score is 0.20 because only the mention of 'xwiki.properties' and OpenID Connect in the 2nd node in retrieval context supports the expected output, while other details like specific property changes, uncommenting instructions, saving, restarting, and automatic activation are not supported by the retrieval context.

CustomContextualRelevancy: The retrieval context mentions xwiki.properties but lacks specific details about setting oidc.enableUser to true for automatic user activation on first login.

Correctness: Actual output incorrectly references `xwiki.cfg` and `xwiki.authentication.createuser=empty` instead of `xwiki.properties` and `oidc.enableUser=true` as in expected output. No mention of user activation, only creation.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.42194240401202115

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0
Correctness: 0.36242365484135775

Reasons:

AnswerRelevancy: The score is 0.77 because the answer provides relevant information about customizing user profile fields in XWiki using OpenID Connect, but the inclusion of multiple source links that do not directly address the question lowers the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'various authentication methods including OpenID Connect, but does not mention customizing user profile fields or mapping them in the `xwiki.properties` file.' Similarly, the second node 'focuses on domain-based authentication and session cookies, which is unrelated to mapping user profile fields from OpenID Connect.' All subsequent nodes continue to lack relevance to the input query about mapping user profile fields from OpenID Connect.

ContextualRecall: The score is 0.40 because while the nodes in the retrieval context mention 'OpenID Connect Authentication' and the 'xwiki.properties' file, they lack specific examples and instructions on mapping properties, replacing placeholders, and applying changes, which are detailed in the expected output.

CustomContextualRelevancy: The retrieval context lacks specific information about customizing user profile fields using `oidc.user.mapping` in `xwiki.properties`, which is detailed in the expected output.

Correctness: The actual output mentions mapping user profile fields but lacks specific instructions on using `xwiki.properties` and `oidc.user.mapping` as the expected output does. It also introduces additional elements like `xwiki.cfg` and listeners not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.572837509107937

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.4444444444444444

CustomContextualRelevancy: 0.5

Correctness: 0.24258061020317698

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.25 because the relevant node, which 'explicitly mentions group synchronization with OpenID Connect, including the use of properties like `oidc.groups.mapping`, is ranked fourth. The first three nodes discuss topics such as 'authentication management and OpenID Connect endpoints', 'container authentication and other authentication methods like LDAP', and 'domain-based authentication and session cookie configuration', none of which are related to group synchronization or OpenID Connect group claims. These irrelevant nodes should be ranked lower than the relevant node.

ContextualRecall: The score is 0.44 because while the 4th node in the retrieval context provides relevant information about configuring group synchronization and mentions properties like

'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', it lacks specific details on locating and editing the 'xwiki.properties' file, replacing placeholders, and saving changes.

CustomContextualRelevancy: The retrieval context mentions group synchronization and the oidc.groups.mapping property, but lacks details on configuring oidc.groups.claim and oidc.groups.separator as specified in the expected output.

Correctness: Actual output lacks specific steps like configuring `oidc.groups.claim` and `oidc.groups.separator`. It introduces new steps not present in expected output, such as using `oidc.userinfoclaims` and custom synchronization through listeners.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.38412370417900604

Individual_scores:

AnswerRelevancy: 0.6842105263157895

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.6205316987582469

Reasons:

AnswerRelevancy: The score is 0.68 because the answer partially addresses the question about file upload restrictions, but it includes multiple irrelevant source links that do not relate directly to configuring these restrictions.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For instance, the first node discusses 'language preferences and translations in XWiki', which does not relate to configuring file upload restrictions. The second node talks about 'various XWiki applications and features', but not about file upload settings. Similarly, the third node covers 'technical details about CSS and XSL for PDF exports', which is also unrelated to the topic of file upload restrictions.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, indicating a complete lack of relevant information.

CustomContextualRelevancy: The retrieval context does not contain any information related to configuring MIME types or setting maximum attachment size as described in the expected output.

Correctness: Actual Output configures specific image MIME types (JPEG, PNG, GIF) instead of any image type, and omits setting max size via UI. It includes restarting the server, which is not in Expected Output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.5439453107179625

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 0.75

ContextualPrecision: 0.5333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.06605281668872733

Reasons:

AnswerRelevancy: The score is 0.71 because the main content addresses the question about refusing script rights to a space administrator, but the presence of multiple irrelevant source link statements lowers the score.

Faithfulness: The score is 0.75 because the actual output incorrectly claims that the 'Script' right is granted to all users by default, while the retrieval context clarifies that its default state is DENY.

ContextualPrecision: The score is 0.53 because relevant nodes are not consistently ranked higher than irrelevant ones. For example, the first node in the retrieval context is about 'language preferences and translations in XWiki', which is unrelated to the input, yet it is ranked higher than relevant nodes.

Similarly, the third node discusses 'interacting with the XWiki RESTful API', which is not relevant to script or admin rights, but is ranked higher than some relevant nodes. However, relevant nodes that discuss 'the implications of various rights, including SCRIPT and ADMIN', and 'details on the Script right and its default settings' are present, contributing to the current score.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the retrieval context, showcasing a complete and accurate match. Great job!

CustomContextualRelevancy: The retrieval context mentions that admin rights include other rights and cannot be denied, but it does not explicitly state that admin rights themselves cannot be refused, which is the key point in the expected output.

Correctness: The actual output provides steps to refuse script rights, contradicting the expected output which states it cannot be refused in XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.4211809681937137

Individual_scores:

AnswerRelevancy: 0.7368421052631579

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.790243703899124

Reasons:

AnswerRelevancy: The score is 0.74 because the answer addresses the main question about configuring team spaces in XWiki, but it is not higher due to repeated irrelevant source links that do not contribute to solving the specific configuration issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question about configuring access permissions in XWiki. For example, the first node discusses 'the structure of an XWiki page, including metadata like author and version', which does not address access control or team spaces. The second node explains 'how to modify resources using XML and HTTP methods', which is not relevant to setting access permissions. Similarly, the third node is about 'adding, editing, and deleting comments', which is unrelated to configuring team access in XWiki.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information about creating groups, setting rights at wiki or space level, or inheriting rights for child pages as mentioned in the expected output.

Correctness: The actual output includes most of the necessary steps such as creating groups, configuring view and edit rights, and testing the configuration. However, it lacks details on setting rights at the wiki level for 'XWikiAllGroup' and inheriting rights to child pages.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.8156800938781242

Individual_scores:

AnswerRelevancy: 0.7058823529411765

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Correctness: 0.7881982103275684

Reasons:

AnswerRelevancy: The score is 0.71 because the answer partially addresses the question about authenticating users with access tokens, but includes multiple irrelevant statements providing source links that do not directly answer the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because relevant nodes in the retrieval context are not consistently ranked higher than irrelevant ones. For instance, the first node discusses 'LDAP authentication settings, which are unrelated to token-based authentication using JWTs,' and should be ranked lower. Similarly, the third node 'describes a method for checking authentication but does not relate to token-based authentication or access tokens,' yet it appears before relevant nodes. However, relevant nodes like the second node, which 'provides information about configuring authorized applications, which is relevant for using access tokens for authentication,' are correctly identified, contributing to the current score.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the nodes in the retrieval context, demonstrating a perfect match. Great job!

CustomContextualRelevancy: The retrieval context includes information about enabling the token authenticator, configuring authorized applications, and generating JWT tokens with required claims and Ed25519 key, aligning closely with the expected output. However, it lacks explicit mention of the limitation regarding tokens for existing users not created through this authenticator.

Correctness: The actual output correctly describes the authentication process using JWT for the LLM Application and matches most facts from the expected output, including the configuration and token requirements. However, it includes additional information on fallback authentication not present in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.28741153771582445

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2244692262949468

Reasons:

AnswerRelevancy: The score is 0.50 because the answer includes multiple irrelevant source links that do not directly address the question about authenticating users with access tokens, which detracts from its overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question about authentication with access tokens. The first node discusses 'model validation strategies in XWiki', which does not relate to authentication. The second node covers 'attachment validation and configuration of mimetypes', also unrelated. The third node provides details on 'the admin UI for model validation', again not pertinent. The fourth node is about 'setting up fields and validation in XWiki', which does not address authentication. Finally, the fifth node describes 'installation steps for a UI in XWiki', which is not connected to the topic of authentication or access tokens.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are central to the expected answer.

Correctness: The actual output mentions token-based authentication but lacks specifics like JWT, required claims, and usage instructions present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.3156164790497742

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.22703220763197868

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the question about authenticating users with access tokens, but includes several irrelevant links that do not directly pertain to the topic, thus preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'upgrading XWiki' and lacks any mention of 'user authentication with access tokens'. The second node focuses on 'features of the Administration Application' without addressing 'authentication with access tokens'. The third node, while mentioning 'user authentication and access authorization', does not specifically cover 'authentication using access tokens'. The fourth node is about 'managing rights and user profiles', which is not related to 'authenticating users with access tokens'. Lastly, the fifth node provides 'a table of rights constants', which is unrelated to the topic of 'authenticating users with access tokens'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output could be attributed to any nodes in the retrieval context, as the context lacks any mention of token-based authentication, JWT specifics, or related configuration details.

CustomContextualRelevancy: The retrieval context lacks any information about token-based authentication, JWT, or the specific configurations mentioned in the expected output.

Correctness: The actual output mentions authentication with access tokens in XWiki and involves configuration, but lacks specific details about JWT requirements, Ed25519 key, and the process outlined in the expected output. It also misses the limitations of the authenticator.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_claude3_5_sonnet

File: qa_001_result.json

Overall_score: 0.3478083348371878

Individual_scores:

AnswerRelevancy: 0.9259259259259259

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.1609240830972009

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and provides useful information on addressing the issue with the bell icon in XWiki. However, it includes a source link that does not directly contribute to solving the problem, which slightly impacts the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'installation instructions and prerequisites', which do not relate to enabling notifications in the wiki. The second node describes 'the notifications application and its features', but does not address enabling notifications. The third node lists 'dependencies for the notifications extension', lacking information on enabling notifications. The fourth node talks about 'overriding notification templates', which is unrelated to enabling notifications in the wiki.

ContextualRecall: The score is 0.00 because the retrieval context does not contain any information about enabling notifications in the wiki or the `notifications.enabled` setting in `xwiki.properties`, which are crucial for the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the `notifications.enabled` setting in `xwiki.properties`.

Correctness: The actual output provides additional installation details irrelevant to enabling notifications via `xwiki.properties`. It omits the specific instruction to set `notifications.enabled` to `true`.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6883072105344747

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.75

ContextualRecall: 0.5714285714285714

CustomContextualRelevancy: 0.5

Correctness: 0.30841469177827696

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the issue of not receiving notifications in XWiki. Great job on providing a focused and helpful answer!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.75 because the first and fourth nodes in the retrieval context are relevant, discussing how to follow a user and various methods of receiving notifications in XWiki, which are crucial for setting up notifications correctly. However, the second node, ranked higher than the fourth, only discusses filters for notifications without addressing enabling notifications or subscriptions. Additionally, the third node, which talks about RSS feeds and notification settings, does not specifically address enabling notifications or subscribing to content, and the fifth node, discussing the 'What's New' application, is not directly related to the issue of enabling notifications.

ContextualRecall: The score is 0.57 because while the nodes in retrieval context cover details about enabling notifications through the alert menu, following users, and checking the network tab (sentences 3, 4, 5, and 6), they do not mention enabling notifications in `xwiki.properties`, subscribing to pages or users, or watching a page or its children (sentences 1, 2, and 3).

CustomContextualRelevancy: The retrieval context covers following users and some notification settings but lacks details on enabling notifications via `xwiki.properties` and watching pages or wikis.

Correctness: The actual output provides guidance on enabling notifications but lacks specific instructions on setting `notifications.enabled` in `xwiki.properties` to `true` and does not mention following users through the User Index. It includes additional information about version-specific settings and advanced filtering, which are not in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9571824315495626

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.9430945892973759

Reasons:

AnswerRelevancy: The score is 0.80 because the output mostly addresses how to enable notifications for one's own actions in XWiki, but it includes irrelevant source and image links that do not directly contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant nodes in the retrieval context are perfectly ranked higher than the irrelevant nodes. The first and second nodes directly address how to enable notifications for one's own actions, mentioning the ability to disable the filter in the 'Advanced filtering options' and switching off the 'Own event filter'. The irrelevant nodes, such as the third node discussing administrative settings and user preferences without addressing the specific question, are correctly

ranked lower. Great job on achieving perfect precision!

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context fully contains the expected answer, including the default behavior of not receiving notifications about one's own activity unless explicitly targeted and the ability to change this in the 'Advanced filtering options' of the notification settings in the user profile.

Correctness: The actual output accurately reflects the expected output by stating that users don't receive notifications about their own activity by default and that the setting can be changed in the 'Advanced filtering options' section of the notification settings. No factual discrepancies or errors are introduced.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6904237839951525

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 0.7

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Correctness: 0.22587603730424927

Reasons:

AnswerRelevancy: The score is 0.92 because the response is mostly relevant and provides useful information on disabling notifications in XWiki. However, it includes a source link and a heading that do not directly address the question, slightly reducing the score.

Faithfulness: The score is 1.00 because there are no contradictions; the actual output perfectly aligns with the retrieval context. Great job!

ContextualPrecision: The score is 0.70 because the first node in the retrieval context is relevant, discussing 'settings' in the notifications menu, which aligns with changing notification settings in the user profile. However, irrelevant nodes, such as the second node discussing 'following a user and default watch settings,' and the third node about 'filtering notifications,' are ranked higher than another relevant node, the fifth node, which mentions 'default auto-watched mode' and user preferences.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the nodes in the retrieval context. Great job on aligning the information perfectly!

CustomContextualRelevancy: The retrieval context mentions default notification settings and user profile settings but lacks specific details about automatic page watching and deleting filters in the 'Custom Filters' list.

Correctness: Actual output does not mention automatic page watching or deleting filters from 'Custom Filters'. It discusses default behaviors in different versions and managing notifications, which diverges from expected focus on automatic watching and filter deletion.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8450151287079777

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 1.0

Correctness: 0.7700907722478658

Reasons:

AnswerRelevancy: The score is 0.90 because the response is largely relevant and helpful in addressing the configuration of OpenID Connect in XWiki. However, it includes a mention of LDAP Authenticator, which is not pertinent to the question about OpenID Connect, preventing a perfect score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node, which mentions the property '# oidc.enableUser=true' directly related to enabling user accounts on first login, is ranked first. This shows a perfect ranking where relevant information is prioritized over irrelevant contexts discussing domain-based authentication, general OpenID Connect information, various authentication methods, and LDAP settings.

ContextualRecall: The score is 0.40 because while the retrieval context's 1st node contains relevant information about enabling user accounts with the property '# oidc.enableUser=true', it lacks details on adding or updating properties, saving the file, restarting XWiki, and the automatic enabling of accounts via OpenID Connect.

CustomContextualRelevancy: The retrieval context contains the exact information found in the expected answer, specifically the mention of the 'oidc.enableUser=true' property in the 'xwiki.properties' file.

Correctness: Actual output correctly identifies the need to set oidc.enableUser=true in xwiki.properties, but lacks specifics about uncommenting the line and restarting XWiki. Provides additional unnecessary information about file location and default settings.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.9259716161731638

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.8

CustomContextualRelevancy: 1.0

Correctness: 0.7558296970389831

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and directly addresses the question about customizing user profile fields in XWiki using OpenID Connect. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which discusses 'oidc.user.mapping=myxproperty1=\${oidc.user.subject}' directly related to mapping additional user profile fields from the OpenID Connect provider to XWiki user properties, is ranked first. This ensures that the most pertinent information is prioritized, resulting in a perfect score. Great job!

ContextualRecall: The score is 0.80 because most of the expected output is well-supported by the 1st node in the retrieval context, particularly the customization and mapping instructions. However, the instruction to restart XWiki after saving changes is not supported by the retrieval context.

CustomContextualRelevancy: The retrieval context contains all the information from the expected output, including the ability to map custom user profile fields using the 'oidc.user.mapping' property in the 'xwiki.properties' file.

Correctness: The actual output correctly describes how to use 'oidc.user.mapping' in 'xwiki.properties', similar to the expected output. It includes examples that match the expected format. However, it introduces additional details about suffixes and custom listeners, which are not mentioned in the expected output but do not contradict it.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.811418661148552

Individual_scores:

AnswerRelevancy: 0.9583333333333334

Faithfulness: 0.9230769230769231

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Correctness: 0.5871017104810554

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant and provides a detailed explanation of configuring group synchronization between XWiki and an OpenID Connect provider. The only minor issue is the inclusion of a source link, which slightly detracts from the directness of the answer.

Faithfulness: The score is 0.92 because the actual output incorrectly claims that the default claim name for groups is 'xwiki_groups', while the retrieval context only mentions that group synchronization is enabled by adding 'xwiki_groups' to the 'oidc.userinfoclaims' property, without specifying it as a default claim name.

ContextualPrecision: The score is 1.00 because all relevant nodes are perfectly ranked higher than irrelevant nodes. The first three nodes in the retrieval context provide detailed information on configuring the 'oidc.groups.claim' and 'oidc.groups.mapping' properties, which are crucial for setting up group synchronization as required. The irrelevant nodes, ranked fourth and fifth, discuss prerequisites, installation instructions, and authentication management, which are not directly related to the specific steps for group synchronization.

ContextualRecall: The score is 0.50 because while nodes in the retrieval context provide detailed information on configuring properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', they lack instructions on locating, opening, and saving the 'xwiki.properties' file, as well as replacing placeholders with actual values.

CustomContextualRelevancy: The retrieval context contains detailed information on configuring group synchronization, including setting the oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping properties, which matches the expected output. However, instructions to locate and edit the xwiki.properties file and restart the XWiki instance are not explicitly mentioned.

Correctness: The actual output includes details on enabling group synchronization and provides examples for configuring group claims and mapping, aligning partially with expected output. However, it omits step-by-step instructions on locating and editing 'xwiki.properties', and lacks information on saving changes and restarting XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6226764970842024

Individual_scores:

AnswerRelevancy: 0.9411764705882353

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.6

CustomContextualRelevancy: 0.3

Correctness: 0.5615491785836456

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant and mostly addresses the input question effectively. However, it could not be higher due to the vague statement 'Additional recommendations:' which does not directly contribute to the solution.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, ranked third, explains how to set the maximum size of an attachment and mentions the XWikiPreferences document, which is relevant to configuring upload size limits. However, irrelevant nodes ranked first and second discuss linking to external files and images, and provide an overview of XWiki's structure and features, which do not address attachment upload restrictions or size limits. These should be ranked lower than the relevant node.

ContextualRecall: The score is 0.60 because while several steps related to configuring attachment properties and maximum upload size in XWiki are well-supported by specific nodes in the retrieval context, some introductory and explanatory sentences do not directly match any content in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific details on configuring 'xwiki.properties' for mimetypes and the exact steps for setting the size to 10MB.

Correctness: The actual output correctly sets the size limit to 10MB and provides steps to configure it, aligning with the expected output. However, it deviates on file type restriction by suggesting the use of an external application instead of configuring the xwiki.properties file directly, missing the specific configuration step.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.6577856181566387

Individual_scores:

AnswerRelevancy: 0.8125

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.13421370893983203

Reasons:

AnswerRelevancy: The score is 0.81 because the main content addresses the question about denying script rights to a space administrator, but it includes irrelevant source links that do not directly address the question, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'programming rights and script execution' but does not address denying script rights to a space administrator. The second node provides a 'table of rights and their default states' without mentioning the inability to deny rights implied by admin rights. The third node explains the 'script right and its availability' but fails to mention denying rights. The fourth node contains 'code snippets related to checking access rights' without addressing the specific issue. Lastly, the fifth node discusses 'converting rights and excluding pages' which is unrelated to the input query.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the nodes in the retrieval context. Great job on ensuring complete alignment!

CustomContextualRelevancy: The expected answer is fully supported by the retrieval context, which states that rights implied by admin cannot be denied, aligning with the expected output.

Correctness: The actual output provides detailed information on managing script rights, but it incorrectly suggests that script rights can be denied to a space administrator, contradicting the expected output that rights implied by admin right cannot be denied.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7658054043883625

Individual_scores:

AnswerRelevancy: 0.8518518518518519

Faithfulness: 0.875

ContextualPrecision: 0.8541666666666666

ContextualRecall: 0.6666666666666666

CustomContextualRelevancy: 0.5

Correctness: 0.8471472411449898

Reasons:

AnswerRelevancy: The score is 0.85 because the response largely addresses the question about configuring permissions in XWiki, but includes several irrelevant links that do not directly assist with the specific configuration task described.

Faithfulness: The score is 0.88 because the claim that 'View' permission will still be inherited from the wiki-level settings contradicts the retrieval context, which states that page-level permissions override wiki-wide permissions.

ContextualPrecision: The score is 0.85 because most relevant nodes are ranked higher than irrelevant ones. The first node discusses 'rights inheritance and administration features' and 'inheritance of access rights,' which are relevant to setting view and edit permissions in XWiki. The second node explains the 'Nested Spaces' concept and rights management, also relevant to configuring space-level rights for teams. However, the third node, which is about string formatting and user identification in OpenID Connect, is not relevant to setting view/edit permissions in XWiki and should be ranked lower. Similarly, the fifth node, which describes the emulation of spaces and pages in XWiki, is not directly related to configuring view/edit permissions for teams and should be ranked lower than the relevant nodes.

ContextualRecall: The score is 0.67 because while several key concepts such as creating groups, setting wiki-wide and space-level rights, and inheriting rights for child pages are supported by nodes in the retrieval context, some introductory and summarizing sentences, as well as the reminder to adjust rights, are not directly referenced in the nodes.

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and inheritance, but lacks specific details about creating groups for each team and setting rights specifically for XWikiAllGroup or individual team spaces.

Correctness: The Actual Output includes all critical details such as setting global view rights, space-specific edit rights, and inheriting rights for child pages. It lacks explicit mention of creating groups for each team, but implies it in the instructions.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.45301965006220085

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8333333333333334

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.2181179003732054

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and addresses the input without any irrelevant statements. Great job!

Faithfulness: The score is 0.83 because the actual output incorrectly attributes the Alert menu to the 'Alerts Application', whereas the retrieval context specifies that the Notifications Application is responsible for this content.

ContextualPrecision: The score is 0.17 because the relevant node in retrieval context, which mentions editing the `xwiki.properties` file to enable notifications, is ranked sixth. The first five nodes in retrieval context are irrelevant as they discuss topics like 'custom notification filters', 'Alert menu', 'different ways to receive notifications', 'various settings and filters', and 'notification email settings', none of which address enabling notifications through the `xwiki.properties` file.

ContextualRecall: The score is 0.50 because while the retrieval context mentions editing the xwiki.properties file (node 5), it lacks the specific instruction to set 'notifications.enabled' to 'true', which is crucial for the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the `xwiki.properties` file or setting `notifications.enabled` to `true`.

Correctness: The actual output does not mention enabling notifications via `notifications.enabled` in `xwiki.properties`, and introduces additional information about applications and permissions that are not in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.33748784278565913

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.14992705671395495

Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses the issue of not receiving notifications, but includes a source link and a comment on search results that are not directly relevant to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant, containing 'image links and references, which are unrelated to the topic of notifications or their settings.' These nodes do not address the issue of enabling notifications, hence they should be ranked lower than any relevant nodes.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context contains image links and references unrelated to the expected answer about enabling notifications and following users or pages in XWiki.

Correctness: The actual output does not mention setting 'notifications.enabled' in 'xwiki.properties', following pages or users, or using the notification menu. It vaguely suggests checking notification settings and email configuration, missing key facts from the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.358646825595176

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2768809535710562

Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses the question about enabling notifications for one's own actions, but it includes a source link that does not directly contribute to solving the issue, slightly reducing the relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customization of page templates, internationalization, accessibility, and other features', which does not address notification settings or user activity. The second node is about 'administrating wiki features like users, groups, and rights', which is unrelated to personal notification settings for user activity. The third node covers 'importing office documents, icon themes, sharing pages, syntax highlighting, and other features', but does not mention notification settings or user activity.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters mentioned in the expected output.

Correctness: The actual output lacks the key detail that notifications for one's own activity are not enabled by default unless explicitly targeted. It suggests navigating to settings to adjust notifications, which aligns partially with the expected output, but doesn't mention 'Erweiterte Filteroptionen'. The advice to consult documentation adds correct additional information.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.3289091503485941

Individual_scores:

AnswerRelevancy: 0.77272727272727

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2007276293642918

Reasons:

AnswerRelevancy: The score is 0.77 because the response partially addresses the question about disabling notifications, but it includes several irrelevant links to XWiki pages that do not pertain to the query, thus preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files', which does not relate to notification settings or disabling notifications. The second node talks about 'moving attachments', which also does not address notification settings. Similarly, the third node is about 'displaying various elements in XWiki', and the fourth node covers 'administrating wiki features', neither of which mention notification settings. The fifth node focuses on 'controlling permissions in the wiki', and the sixth node lists 'various applications and extensions for XWiki', none of which are relevant to the input query about disabling notifications.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to automatic page following or notification settings as mentioned in the expected output.

Correctness: The actual output does not mention 'Automatisches Verfolgen von Seiten' or 'Benutzerdefinierte Filter'. It inaccurately focuses on deactivating notifications and sources not related to the expected content.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.5381401830334357

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.75

CustomContextualRelevancy: 0.0

Correctness: 0.2288410982006143

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and directly addresses the input question without any irrelevant statements. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.25 because the relevant node in the retrieval context, which discusses updating the 'xwiki.properties' file to enable automatic user activation, is ranked fourth. The first node discusses 'domain-based authentication and session cookie configuration', which is unrelated to the specific configuration needed. The second node provides 'general information about OpenID Connect authentication and configuration guides', which does not address the specific need. The third node provides 'metadata about the OpenID Connect Authenticator', which is also not relevant to the configuration needed for automatic user activation.

ContextualRecall: The score is 0.75 because the 4th node in the retrieval context provides relevant information about modifying 'xwiki.properties' and setting 'oidc.enableUser' to true, which supports the expected output. However, there is no direct mention in the nodes about ensuring new user accounts are automatically activated upon first login, which affects the completeness of the retrieval.

CustomContextualRelevancy: The retrieval context does not mention adjusting the 'xwiki.properties' file or the 'oidc.enableUser' property, which are key elements in the expected output.

Correctness: Actual output suggests different properties like 'oidc.user.autocreate' instead of 'oidc.enableUser' and lacks specific instruction to uncomment and set to 'true'. It provides additional irrelevant steps and lacks focus on the required property modification.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.536521736600025

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.8

CustomContextualRelevancy: 0.0

Correctness: 0.49055899102872136

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and mostly addresses the user's question about customizing user profile fields in XWiki. However, it slightly falls short of a perfect score due to a statement about the lack of details in the provided documents, which is not directly related to the user's query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect endpoints' without mentioning user profile field mapping. The second node is about 'container authentication and other authentication methods like LDAP and Active Directory', which do not relate to customizing OpenID Connect user profile fields. The third node provides 'general information about OpenID Connect authentication and configuration guides' but does not address mapping user profile fields. The fourth node focuses on 'OpenID Connect Authenticator extension details' without mentioning user profile field customization. Lastly, the fifth node lists 'dependencies for the OpenID Connect Authenticator extension', which is unrelated to user profile field customization.

ContextualRecall: The score is 0.80 because the nodes in retrieval context support the general process of configuring user profile fields in 'xwiki.properties' and using 'oidc.user.mapping', but they do not explicitly cover the specific example of replacing 'myxproperty1', 'myxproperty2', or '{oidc.user.subject}'.

CustomContextualRelevancy: The retrieval context does not contain any information about customizing user profile fields using 'xwiki.properties' or 'oidc.user.mapping', which are key elements of the

expected output.

Correctness: The actual output partially aligns with the expected output by mentioning configuration in `xwiki.properties` and mapping rules, but lacks specific examples and details provided in the expected output, such as the use of variables like `\${oidc.user.subject}`. It also suggests consulting external resources, which is not in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.3690191310939823

Individual_scores:

AnswerRelevancy: 0.9565217391304348

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.25759304743345923

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant and provides a clear explanation of the setup process, with only a minor mention of lacking specific configuration details that slightly detracts from a perfect score.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and endpoints for OpenID Connect' but lacks information on 'group synchronization or configuration in xwiki.properties'. The second node covers 'container authentication and other methods like LDAP', missing details on 'group synchronization or specific properties needed for OpenID Connect group claims'. The third node mentions 'OpenID Connect and automatic synchronization' but lacks specifics on 'configuring group synchronization in XWiki using xwiki.properties'. The fourth node lists 'dependencies for the OpenID Connect extension' without relating to 'configuration of group synchronization'. Lastly, the fifth node provides 'general information about the OpenID Connect authenticator' without addressing 'group synchronization setup'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, indicating a complete lack of relevant information in the context provided.

CustomContextualRelevancy: The retrieval context lacks details on configuring group synchronization properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', which are essential in the expected output.

Correctness: The actual output lacks specific instructions for 'oidc.groups.separator' and 'oidc.groups.mapping' configuration, and does not mention restarting XWiki to apply changes, deviating from critical steps in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6470413371578836

Individual_scores:

AnswerRelevancy: 0.8
Faithfulness: 1.0
ContextualPrecision: 0.325
ContextualRecall: 0.7
CustomContextualRelevancy: 0.5
Correctness: 0.5572480229473011

Reasons:

AnswerRelevancy: The score is 0.80 because the response mostly addresses the input question about upload restrictions, but it includes irrelevant source links that do not directly relate to the question, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant nodes in the retrieval context, specifically the fourth node mentioning 'Size Limit' and 'Mimetype Restriction', and the fifth node providing steps on setting maximum attachment size, are ranked lower than irrelevant nodes. The first node discusses 'trusted domains and PDF export customization', the second node is about 'ODT and PDF export features', and the third node lists 'various macros and features', none of which relate to the input about upload restrictions. These irrelevant nodes should have been ranked lower to improve the score.

ContextualRecall: The score is 0.70 because while the 5th node in the retrieval context supports the steps for configuring the maximum attachment size (as detailed in sentences 5-10 of the expected output), it lacks information on configuring MIME types and editing the `xwiki.properties` file (as described in sentences 1-4).

CustomContextualRelevancy: The retrieval context mentions setting limits on attachment size and MIME type restrictions, but lacks specific details on configuring MIME types in xwiki.properties and setting the size to 10 MB as outlined in the expected output.

Correctness: The actual output correctly describes setting maximum upload size and emphasizes separate configuration for each (sub)wiki. However, it suggests using the Attachment Validation Application instead of updating 'xwiki.properties' for MIME type restrictions, which is a key discrepancy from the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_019_result.json

Overall_score: 0.3441348257091412
Individual_scores:
AnswerRelevancy: 0.7692307692307693
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.29557818502407807

Reasons:

AnswerRelevancy: The score is 0.77 because the response partially addresses the question about denying script rights to a space administrator, but includes irrelevant source links and lacks specific information directly answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'form validation and JavaScript', the second node is about 'template providers and page creation', and the third node involves 'custom configuration and class references'.

None of these topics relate to script rights or administrative permissions, which is the focus of the input.
ContextualRecall: The score is 0.00 because the expected output sentence about the limitation of XWiki regarding admin rights is not supported by any information in the nodes in the retrieval context.
CustomContextualRelevancy: The retrieval context does not contain any information regarding the support or denial of admin rights in XWiki, which is the core of the expected answer.
Correctness: The actual output suggests a method to deny script rights, which contradicts the expected output stating it is not supported. It includes extra steps and cautions not present in the expected output.
Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_020_result.json

Overall_score: 0.49213561867370315
Individual_scores:
AnswerRelevancy: 0.8275862068965517
Faithfulness: 1.0
ContextualPrecision: 0.25
ContextualRecall: 0.14285714285714285
CustomContextualRelevancy: 0.0
Correctness: 0.7323703622885246
Reasons:

AnswerRelevancy: The score is 0.83 because the answer mostly addresses the configuration question in XWiki, but the presence of multiple source links that do not directly contribute to solving the problem slightly detracts from its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.25 because only the fourth node in the retrieval context is relevant, as it provides a link to managing users and groups, which is pertinent to creating groups for each team and setting user rights. However, the first three nodes are ranked higher and are irrelevant, discussing topics like XWiki's database efficiency, notifications, and nested pages, which do not relate to configuring access controls for teams. Additionally, the fifth node, also irrelevant, covers group synchronization and customization, not directly addressing the configuration of access controls for teams in XWiki.

ContextualRecall: The score is 0.14 because only the sentence about creating groups for each team (sentence 1) is supported by node 4 in the retrieval context, while all other sentences in the expected output lack corresponding information in the retrieval context nodes.

CustomContextualRelevancy: The retrieval context lacks any information related to setting access controls, creating groups, or managing rights in XWiki as described in the expected output.

Correctness: The actual output accurately reflects group creation and setting of 'View' and 'Edit' rights but includes additional 'Delete' rights not present in the expected output. The recursive rights inheritance is correctly mentioned.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_021_result.json

Overall_score: 0.35257783601612286
Individual_scores:
AnswerRelevancy: 0.9375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.17796701609673685

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant to the issue of the notification bell not appearing, with only a minor irrelevant detail about XWiki syntax documentation that slightly detracts from the focus.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'custom notification filters and their behavior', the second node lists 'various notification-related links and settings', the third node provides 'links to notification preferences and macros', the fourth node describes 'the notification application and its features', and the fifth node discusses 'obtaining an RSS feed for notifications and settings'. None of these nodes mention 'enabling notifications in the wiki settings', which is relevant to the input.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications through the `xwiki.properties` file or setting `notifications.enabled` to `true`, which is required in the expected output.

Correctness: The actual output does not address the activation of notifications via `xwiki.properties` as mentioned in the expected output. Instead, it provides troubleshooting steps for visible alerts and extensions, which deviates from the expected content.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.6685772458102469

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Correctness: 0.2947968081948155

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant to the issue of not receiving notifications, but it includes a source link that does not directly address the problem, slightly reducing its relevancy.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job on maintaining perfect alignment!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than irrelevant nodes. The first node discusses 'Enable and disable the extension' and 'Enabling notifications for all users by default', which directly relates to ensuring notifications are activated. The second node includes 'Follow a user', aligning with the expected output's mention of following users to receive notifications. The irrelevant nodes, ranked third and fourth, discuss 'RSS feeds and macro settings' and 'various notification preferences macros', which do not directly address enabling notifications or following users.

ContextualRecall: The score is 0.50 because while some elements like enabling notifications and following users or pages are supported by nodes in the retrieval context, specific details such as setting 'notifications.enabled' in 'xwiki.properties', the 'alerte' menu, receiving all notifications from a user, and the 'réseau' tab are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' to true and specific steps for following pages or users.

Correctness: The actual output lacks specific instructions about setting 'notifications.enabled' to 'true' in 'xwiki.properties' and following pages or users. It discusses notification settings and filters but misses key steps from the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8269174473745928

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Correctness: 0.7281713509142237

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the question about enabling notifications for one's own actions, but it includes irrelevant source links that do not directly relate to the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which states, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' is ranked first. This directly addresses the question about enabling notifications for one's own actions, while the irrelevant nodes, which discuss topics like RSS feeds, general notification settings, and technical aspects without addressing the specific question, are ranked lower. Great job on getting it spot on!

ContextualRecall: The score is 0.50 because the first sentence of the expected output is supported by the 1st node in the retrieval context, but the second sentence lacks support as there is no mention of 'Options de filtrage avancées' or similar settings in the nodes in retrieval context.

CustomContextualRelevancy: The retrieval context confirms that by default, users do not receive notifications for their own actions and mentions the ability to change this in settings, aligning closely with the expected output. However, it does not specifically mention 'Options de filtrage avancées'.

Correctness: The actual output correctly states that notifications for one's own actions are not received by default and can be changed by disabling a filter. However, it introduces unnecessary steps for accessing settings and lacks mention of 'Options de filtrage avancées' as in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7832995711127703

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Correctness: 0.26090853778773193

Reasons:

AnswerRelevancy: The score is 0.89 because the response mostly addresses the question about disabling notifications, but includes irrelevant links that do not directly help with managing notifications.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant node. The first three nodes effectively address the input by discussing topics like 'Choosing the default auto-watched mode', 'Create global notification filters', 'Notifications Auto Watch Preferences', and 'Default Watch Settings', which are all pertinent to modifying notification settings. The irrelevant node, which focuses on 'getting a notification RSS feed and general notification settings', is appropriately ranked last as it does not directly address the issue of disabling automatic notifications.

ContextualRecall: The score is 0.75 because most sentences in the expected output are supported by the 3rd node in the retrieval context, which discusses 'auto watched pages', 'Notifications Auto Watch Preferences', and 'custom notification filters'. However, the retrieval context does not mention being notified about changes to your work, which affects the score.

CustomContextualRelevancy: The retrieval context includes information about auto-watched pages and custom filters, which aligns with the expected answer's mention of automatic page following and notification filters. However, it lacks explicit details on modifying the 'Suivi automatique des pages' setting and removing filters in 'Filtres personnalisés'.

Correctness: The actual output provides some general information about managing notifications but lacks specific details about automatic page tracking and modifying filters in 'Filtres personnalisés' as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.42386990926310464

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.20988612224529443

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to configuring OpenID Connect for automatic user account activation, but it includes an irrelevant mention of the LDAP Authenticator, which slightly detracts from the focus.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based instance authentication and bypassing OpenID Connect',

which does not address automatic user activation. The second node mentions 'container authentication and various authentication methods', but not the automatic activation with OpenID Connect. The third node provides 'OpenID Connect authentication and configuration guides', yet omits the specific property `oidc.enableUser` for automatic activation. The fourth node lists 'dependencies for the OpenID Connect extension', lacking any discussion on user account activation. Lastly, the fifth node focuses on 'LDAP and Active Directory integration', which is unrelated to the input query.

ContextualRecall: The score is 0.40 because while the nodes in the retrieval context mention 'OpenID Connect Authentication' and 'xwiki.properties', which relate to adjusting the file for OpenID Connect in XWiki (supporting sentence 1), they do not mention adding or updating specific properties, saving the file, restarting XWiki, or automatic activation of new user accounts (unsupporting sentences 2, 3, and 4).

CustomContextualRelevancy: The retrieval context does not mention adjusting xwiki.properties or setting oidc.enableUser to true for automatic user activation, which are key points in the expected output.

Correctness: The actual output mentions the wrong configuration file (`xwiki.cfg`) and property (`xwiki.authentication.createuser=empty`) which do not align with the expected output's instructions for `xwiki.properties` and `oidc.enableUser=true`. Additionally, it claims no specific information is available, which is factually incorrect. Completeness and accuracy are lacking.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.44490976106513275

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.26945856639079657

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and directly addresses the question about customizing user profile fields from an OpenID Connect provider in XWiki. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect provider configuration', but it does not mention 'user profile field mapping or customization'. The second node is about 'container authentication and LDAP', which also lacks information on 'customizing user profile fields or mapping'. The third node provides 'general information about the OpenID Connect Authenticator extension', but does not address 'user profile field mapping or customization'. The fourth node lists 'dependencies for the OpenID Connect Authenticator extension', which is unrelated to 'user profile field customization or mapping'. Finally, the fifth node describes 'configuration guides for various systems', but does not include information on 'mapping or customizing user profile fields in XWiki'.

ContextualRecall: The score is 0.40 because while the 1st node in the retrieval context supports the general concept of customizing user profile fields using 'xwiki.properties' and 'oidc.user.mapping', it lacks specific examples, instructions on replacing property names, and details about saving the file and restarting XWiki as mentioned in the expected output.

CustomContextualRelevancy: The retrieval context does not contain information about customizing user profile fields using `oidc.user.mapping` in `xwiki.properties` or the process of saving and restarting

XWiki.

Correctness: The actual output suggests using `xwiki.properties` for mapping but lacks specific details like `myxproperty1=\${oidc.user.subject}` and the need to restart XWiki. It introduces extra steps like checking logs and contacting the community.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5351236947070299

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.3076923076923077

CustomContextualRelevancy: 0.3

Correctness: 0.2697165272165384

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider, with no irrelevant information included. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which is the third node, provides information on group synchronization and mentions the use of claims and the `oidc.groups.mapping` property. However, the first, second, fourth, and fifth nodes are ranked higher or close to the relevant node and focus on topics like 'authentication management and endpoints configuration', 'container authentication and LDAP', 'general information about OpenID Connect authentication', and 'dependencies for the OpenID Connect extension', which are not directly related to group synchronization.

ContextualRecall: The score is 0.31 because while the 3rd node in the retrieval context supports the general concept of configuring group synchronization in XWiki, it lacks specific details such as locating and editing the `xwiki.properties` file, providing examples, and instructions for saving and restarting XWiki, which are crucial for a complete understanding.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using `oidc.groups.mapping`, but lacks details on configuring `oidc.groups.claim` and `oidc.groups.separator`, which are present in the expected output.

Correctness: The actual output lacks key instructions such as setting `oidc.groups.claim` and `oidc.groups.separator`. It introduces `oidc.userinfoclaims` and mentions event listeners, which are not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.384780199148768

Individual_scores:

AnswerRelevancy: 0.7727272727272727

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.5359539221653353

Reasons:

AnswerRelevancy: The score is 0.77 because the answer provides relevant information about configuring file upload restrictions, but it includes multiple source links that do not directly address the question, reducing the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'language preferences and translations in XWiki', which does not relate to configuring file upload restrictions. The second node talks about 'various XWiki applications and their functionalities', again not mentioning file upload settings. The third node is about 'CSS and XSLT processing for PDF exports', which is unrelated to the topic. The fourth node lists 'various macros and their descriptions in XWiki', none of which pertain to file upload settings. The fifth and sixth nodes contain 'links to image files and references', which are not relevant to file upload restrictions. Hence, irrelevant nodes are ranked higher than relevant ones, resulting in a score of 0.00.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about configuring MIME types or setting maximum attachment size as outlined in the expected output.

Correctness: The actual output matches the expected output in terms of setting MIME types and max file size, but lacks specific steps such as editing 'xwiki.properties' and using the 'XWikiPreferences' editor, and introduces additional instructions like testing the upload and informing users.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.7074610086890386

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.7947660521342315

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the question but includes irrelevant source links that do not directly relate to refusing script rights to a space administrator.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining accuracy and alignment with the retrieval context!

ContextualPrecision: The score is 0.58 because relevant nodes, like the second node explaining that 'programming rights, which include script rights, are tied to administrative permissions and cannot be denied separately,' and the third node mentioning 'administrative rights imply script rights,' are ranked after the first node. The first node is irrelevant as it discusses 'language preferences and translations in XWiki,' which is unrelated to script rights. Additionally, the fourth and fifth nodes, which discuss 'checking access rights programmatically' and 'executing tasks with specific user rights,' are also irrelevant and ranked after relevant nodes.

ContextualRecall: The score is 1.00 because every part of the expected output is perfectly aligned with the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context does not explicitly mention that admin rights cannot be denied, which is a key point in the expected output. It only lists rights and their properties without addressing the specific case of admin rights being non-deniable.

Correctness: The actual output correctly explains that administrative rights, including 'SCRIPT', cannot be individually denied, aligning with the expected output that such rights cannot be refused. However, it adds additional information about rights inheritance not explicitly mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.40688061091112554

Individual_scores:

AnswerRelevancy: 0.8260869565217391

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.6151967089450137

Reasons:

AnswerRelevancy: The score is 0.83 because the main content effectively addresses how to configure access rights in XWiki, but it includes multiple irrelevant source links that do not contribute to the solution.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'the structure of an XWiki page', which does not relate to configuring access rights for teams. The second node talks about 'modifying resources via XML and HTTP methods', which is not relevant to setting access permissions in XWiki. The third node is about 'editing comments and attachments', which is unrelated to configuring team access rights in XWiki. The fourth node explains 'the creation of spaces and pages in XWiki', but does not provide information on setting access rights for teams. The fifth node is about 'actions needed after running a migrator', which does not pertain to configuring team access rights in XWiki. Finally, the sixth node discusses 'configuration display scopes in XWiki', which is not directly related to setting team-specific access rights.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any node in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting rights at the wiki or space level, or inheriting rights for child pages as outlined in the expected output.

Correctness: The actual output covers the major steps like configuring global view rights and space-specific edit rights. However, it misses specific instructions on 'Hériter des droits pour les pages enfants' and doesn't mention setting 'Afficher' for 'XWikiAllGroup'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7650360667931605

Individual_scores:

AnswerRelevancy: 0.7619047619047619

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Correctness: 0.8140259245684864

Reasons:

AnswerRelevancy: The score is 0.76 because the output mostly addresses the question about authenticating users with access tokens, but it includes multiple source links and a general introductory phrase that do not directly contribute to answering the question, thus preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because some relevant nodes are ranked lower than irrelevant ones. The first node is about 'LDAP configuration' and not about 'token-based authentication', yet it is ranked higher than the second node, which is relevant as it 'discusses the configuration of authorized applications'. Similarly, the third node focuses on 'OpenID Connect scopes', which is not directly related, but is ranked higher than the fourth node that 'describes the JWT token authenticator', which is directly relevant. This mix of rankings results in a balanced score.

ContextualRecall: The score is 0.71 because while several key aspects of the expected output, such as the use of JWTs and configuration steps, are supported by nodes in the retrieval context, there are missing details regarding generating a JWT token with specific claims and including the token in the 'Authorization' header, which are crucial for a complete understanding.

CustomContextualRelevancy: The retrieval context includes information about enabling the token authenticator, configuring authorized applications, and using JWTs, which aligns with the expected answer. However, details about specific JWT claims and the use of Ed25519 keys are not explicitly mentioned.

Correctness: The Actual Output correctly identifies the use of JWT for authentication, the need to set `xwiki.authentication.authclass`, and configuring 'Authorized Applications'. However, it inaccurately mentions a fallback authenticator which isn't in the Expected Output, and misses details on specific claims required in the JWT.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.33596592933170233

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.26579557599021386

Reasons:

AnswerRelevancy: The score is 0.75 because the response addresses the question about authenticating users with access tokens, but includes irrelevant source links and lacks specific information directly related to the topic.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'model validation strategies in XWiki', the second node is about 'attachment validation and configuration in XWiki', and the third node provides information on 'the admin UI for model validation'. None of these are related to authenticating users with access tokens, resulting in a score of 0.00.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are central to the expected answer.

Correctness: The actual output mentions token-based authentication and the Authorization header format, but lacks specifics on JWT, Ed25519 signature, and configuration details in xwiki.properties and authorized applications as outlined in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.34221910193647714

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.17831461161886303

Reasons:

AnswerRelevancy: The score is 0.88 because the answer effectively addresses the question about authenticating users with access tokens, but it includes source links that are not directly related to the topic, slightly reducing its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'upgrading XWiki' and lacks any mention of 'user authentication or access tokens'. The second node talks about 'features provided by the Administration Application' but does not address 'authentication with access tokens'. The third node covers 'rights management and user profiles' without mentioning 'authentication with access tokens or JWTs'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context. The retrieval context lacks any mention of authentication with token-based extensions, JWT, Ed25519 keys, or instructions for activating authenticators and configuring applications.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication or JWT, which are central to the expected output.

Correctness: The actual output mentions JWT but lacks specific setup instructions like setting `xwiki.authentication.authclass` and configuring authorized applications, which are key facts in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_command-r_35B_Q4

File: qa_001_result.json

Overall_score: 0.30532534932314787

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.03195209593888694

Reasons:

AnswerRelevancy: The score is 0.80 because the response mostly addresses the issue of fixing the 'Bell' icon in XWiki, but includes irrelevant links that do not directly help solve the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question about enabling notifications in XWiki. The first node is about 'installation instructions and prerequisites for extensions', which does not help with the issue. The second node describes 'the functionality of the Alert menu and notifications' without addressing the enabling process. The third node lists 'dependencies related to the notifications extension', but again, does not solve the problem. Finally, the fourth node discusses 'overriding notification templates', which is unrelated to the query about enabling notifications.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be linked to any nodes in the retrieval context, as there is no mention of enabling notifications or the `xwiki.properties` settings.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the `notifications.enabled` setting in `xwiki.properties`, which is the key information in the expected output.

Correctness: The actual output does not address enabling notifications via 'notifications.enabled' in 'xwiki.properties' as stated in the expected output. Instead, it discusses issues with the 'Bell' icon and suggests installing the 'Alerts Application', which is unrelated to the expected task.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6321514787801626

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 0.9090909090909091

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Correctness: 0.251465022413596

Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses the issue of not receiving notifications in XWiki, but includes irrelevant source links that do not directly contribute to solving the problem.

Faithfulness: The score is 0.91 because the actual output incorrectly claims that no application types are enabled for notifications in XWiki versions less than 15.5, while the retrieval context specifies that certain types, such as Mention, are indeed enabled.

ContextualPrecision: The score is 0.75 because the first and fourth nodes in the retrieval context are relevant, explaining how to follow a user and listing ways to receive notifications, which directly address

the issue of setting up notifications in XWiki. However, the second node, ranked higher than the fourth, discusses filters and settings without directly addressing enabling notifications, and the third node, also ranked higher than the fourth, mentions RSS feeds without providing information on enabling notifications. Additionally, the fifth node discusses the 'What's New' application, which is unrelated to managing notifications, contributing to the score not being higher.

ContextualRecall: The score is 0.50 because while some details about receiving notifications and following users are supported by nodes in the retrieval context, key instructions such as enabling notifications in `xwiki.properties` and subscribing to pages are missing from the context.

CustomContextualRelevancy: The retrieval context includes information about following users and receiving notifications, but lacks details on enabling notifications via `xwiki.properties` and watching pages or wikis.

Correctness: The actual output covers enabling notifications but includes unrelated details about filters, browser settings, and version discrepancies not present in the expected output. It lacks specific instructions on watching pages or users.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9348777766659011

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.859266659995407

Reasons:

AnswerRelevancy: The score is 0.75 because the response mostly addresses the question about enabling notifications for one's own actions in XWiki, but includes irrelevant source links that do not directly contribute to solving the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Keep up the great work!

ContextualPrecision: The score is 1.00 because the relevant nodes in the retrieval context, which explain that 'By default, one don't receive notification about one's own activity' and provide solutions to 'disable this filter in the "Advanced filtering options" section of your own notification settings' (first node), are ranked higher than irrelevant nodes. The irrelevant nodes, such as the one discussing 'administrators configuring notification settings for users' (third node), are correctly ranked lower. Great job on the precision!

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context contains all the information from the expected answer, including the default behavior of not receiving notifications about one's own activity unless explicitly targeted, and the ability to disable this filter in the 'Advanced filtering options' section of the notification settings in the user profile.

Correctness: The actual output correctly states the default behavior and how to disable it, matching the expected output. It provides additional detail on navigating to the setting, which doesn't contradict the expected output but enhances user understanding.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5644792278266976

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Correctness: 0.2118753669601857

Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses how to disable unwanted notifications in XWiki, but includes a statement about default notification settings, which is not directly relevant to the user's question.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job maintaining perfect alignment!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'getting a notification RSS feed and general notification settings' but does not address disabling notifications or automatic page watching. The second node talks about 'following a user and default watch settings' without mentioning how to disable notifications. The third node mentions 'filters for notifications' but lacks information on disabling notifications. The fourth node discusses 'enabling and disabling email notifications by an administrator,' which is not directly related to the user's concern. The fifth node is about 'administrators configuring notification settings and default auto-watched modes,' but it does not explain how a user can disable notifications. The sixth node is about 'overriding default notification templates and installation instructions,' which is not relevant to the user's query.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context mentions default notification settings and user profile settings but lacks details on automatic page watching and deleting notification filters.

Correctness: The actual output does not mention automatic page watching after major modifications or how to change this setting in the profile, and introduces new details about following users and events, which are unsupported.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.799220159303947

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 1.0

Correctness: 0.5619876224903482

Reasons:

AnswerRelevancy: The score is 0.83 because the answer mostly addresses the question about configuring XWiki for OpenID Connect, but includes some irrelevant details about LDAP Authenticator and skipping authentication, which are not directly related to enabling user accounts upon first login.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions the property '# oidc.enableUser=true' directly related to enabling user accounts on first login, is ranked first. This ensures that the relevant information is prioritized over irrelevant nodes discussing topics like domain-based instance configuration, general OpenID Connect information, container authentication, and LDAP configuration.

ContextualRecall: The score is 0.40 because while the 1st node in retrieval context supports the mention of the 'oidc.enableUser' property, it lacks explicit guidance on adding or updating properties, saving the file, and restarting the XWiki instance, as well as the outcome of automatically enabling user accounts upon first login.

CustomContextualRelevancy: The retrieval context contains the exact information found in the expected answer, specifically the property '# oidc.enableUser=true' which needs to be uncommented and set to 'true' in the xwiki.properties file.

Correctness: The actual output correctly mentions setting the 'oidc.enableUser' to 'true', but it incorrectly refers to 'xwiki.cfg' instead of 'xwiki.properties' and adds unnecessary details about skipping authentication which aren't in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.8615847382953193

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 1.0

Correctness: 0.7604175206810069

Reasons:

AnswerRelevancy: The score is 0.91 because the response effectively addresses the question about customizing user profile fields in XWiki with OpenID Connect, but includes a source link that doesn't directly contribute to the solution.

Faithfulness: The score is 1.00 because there are no contradictions; the actual output perfectly aligns with the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context is ranked first. It mentions 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user' and provides examples of how to map properties using 'oidc.user.mapping', which directly relates to customizing user profile fields. Great job on getting it spot on!

ContextualRecall: The score is 0.50 because while the retrieval context provides information on mapping non-standard properties from the OpenID Connect provider to XWiki user properties (node 3), it lacks details on replacing placeholders with actual property names and the necessity of saving and restarting XWiki to apply changes.

CustomContextualRelevancy: The retrieval context contains all the information from the expected answer, including the ability to customize user profile fields using the 'oidc.user.mapping' property in the 'xwiki.properties' file, with examples provided.

Correctness: The actual output aligns with the expected output regarding the ability to customize user profile fields and the use of 'oidc.user.mapping' in 'xwiki.properties'. However, the example mappings differ; the expected uses '\${oidc.user.subject}' while the actual uses '\${oidc.user.given_name}',

`\${oidc.user.family_name}`, and `\${oidc.user.email}`. The sequence and logic are consistent, but the specific variable examples differ.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.7316034739687494

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 0.6666666666666666

ContextualPrecision: 1.0

ContextualRecall: 0.45454545454545453

CustomContextualRelevancy: 0.9

Correctness: 0.5684087226003744

Reasons:

AnswerRelevancy: The score is 0.80 because the response mostly addresses the input question about configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes some irrelevant statements that do not contribute to the explanation, preventing a higher score.

Faithfulness: The score is 0.67 because the actual output incorrectly states that adding the claim `xwiki_groups` to `oidc.userinfoclaims` is the default setting, which is not specified in the retrieval context. Additionally, the actual output refers to the `oidc.groups.separator` parameter for parsing groups, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant ones. Great job on getting everything in the right order!

ContextualRecall: The score is 0.45 because while some details about configuring group synchronization, such as claims and separators, are supported by nodes in the retrieval context, many procedural steps and introductory statements in the expected output are not explicitly covered by the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context includes detailed instructions on configuring group synchronization using properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, matching the expected output. However, it lacks explicit mention of saving the file and restarting the XWiki instance.

Correctness: The actual output correctly mentions configuring `oidc.groups.mapping` and `oidc.groups.separator`, similar to the expected output, but lacks detail on `oidc.groups.claim` setup and saving changes. It introduces `oidc.groups.allowed` and `oidc.groups.forbidden`, which are not in the expected output, indicating discrepancies.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.5370942935287676

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 0.8

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.8888888888888888

CustomContextualRelevancy: 0.2

Correctness: 0.12534353895038342

Reasons:

AnswerRelevancy: The score is 0.88 because the response largely addresses the question about setting upload restrictions, but it includes a link to basic concepts of XWiki, which is not directly relevant to the specific query.

Faithfulness: The score is 0.80 because the actual output incorrectly claims that the documentation lacks details on setting a size limit for uploads, while the retrieval context clearly states that administrators can set limits on the maximum size of an attachment in XWiki.

ContextualPrecision: The score is 0.33 because the relevant node, ranked third, provides information about 'setting the maximum size of an attachment in XWikiPreferences', which directly addresses the input query. However, irrelevant nodes are ranked higher, such as the first node discussing 'linking to external files and backlinks indexation', and the second node offering 'general information about XWiki's structure and features', neither of which relate to configuring upload restrictions or size limits.

ContextualRecall: The score is 0.89 because while the 3rd node in the retrieval context provides detailed instructions for configuring mimetypes and setting maximum attachment size, the overall goal of ensuring only images or PDF files of max 10MB can be uploaded is not explicitly stated in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific details on configuring mimetypes and setting size to 10MB as outlined in the expected output.

Correctness: The actual output lacks specific steps to configure mimetype restrictions and max attachment size as detailed in the expected output, and suggests consulting documentation without addressing the task directly.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.569527885265984

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 0.6666666666666666

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Correctness: 0.041409735838327646

Reasons:

AnswerRelevancy: The score is 0.91 because the response is mostly relevant and provides useful information on the topic, but it includes a link about content organization and nested pages migration, which is not directly related to denying script rights to a space administrator.

Faithfulness: The score is 0.67 because the actual output incorrectly claims that the 'Script' right is allowed for all users at the main wiki level and that all users can execute scripts unless revoked, whereas the retrieval context clarifies that this right is not granted by default in XWiki 14.10+.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses programming rights but does not address denying script rights to a space administrator. The second node lists various rights without mentioning the inability to deny script rights. The third node explains script rights but not their denial. The fourth node provides technical details unrelated to denying script rights. The fifth node discusses rights conversion during migration without mentioning script rights denial.

ContextualRecall: The score is 1.00 because the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context confirms that rights implied by admin cannot be denied, aligning with the expected output. However, it does not explicitly mention the lack of support for denying implied rights in XWiki, leading to a slight deduction.

Correctness: The actual output does not address the key fact that rights implied by admin right cannot be denied in XWiki. It focuses on script rights management instead, which is not relevant to the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.645520425590486

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 0.7142857142857143

ContextualPrecision: 0.29166666666666663

ContextualRecall: 0.875

CustomContextualRelevancy: 0.5

Correctness: 0.5974333304852715

Reasons:

AnswerRelevancy: The score is 0.89 because the answer mostly addresses the question about configuring permissions in XWiki, but it includes irrelevant lists of sources that do not directly help with the configuration task.

Faithfulness: The score is 0.71 because the actual output inaccurately claims that you can set rights for each group on their corresponding space and access 'Space Administration', while the retrieval context indicates that the concept of space has been removed from the XWiki UI.

ContextualPrecision: The score is 0.29 because the relevant nodes in the retrieval context, specifically the fourth node which 'provides information on XWiki's permission system, including the ability to set view and edit rights at different levels,' and the sixth node which 'outlines the rules for setting rights in XWiki, including wiki-wide and page-level permissions,' are ranked lower than irrelevant nodes. The first node, which 'discusses the historical organization of content in XWiki and the concept of Nested Pages,' and the second node, which 'explains the parent/child relationship and the introduction of Nested Pages,' are ranked higher despite not addressing the configuration of access rights for teams. This misordering of relevant and irrelevant nodes results in a lower score.

ContextualRecall: The score is 0.88 because most of the expected output sentences are supported by the nodes in the retrieval context, such as setting rights at different levels and creating groups for each team. However, there is a lack of specific mention in the nodes regarding adjusting rights for additional teams or spaces, which slightly lowers the score.

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and inheritance, but lacks specific details about creating groups for each team and setting rights for XWikiAllGroup as mentioned in the expected output.

Correctness: The actual output aligns with expected output on creating groups and setting edit rights, but lacks clarity on setting view rights for 'XWikiAllGroup' and ensuring inheritance for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.47458766843786265

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.18085934396050934

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant to the input, with no irrelevant statements present. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node in the retrieval context, which mentions editing the 'xwiki.properties' file, is ranked sixth. This is much lower than the irrelevant nodes, such as the first node discussing 'custom notification filters and their behavior', and the second node describing 'the alert menu and notification display', which do not relate to enabling notifications in XWiki. The relevant information should be ranked higher to improve the score.

ContextualRecall: The score is 0.50 because while the 6th node in the retrieval context discusses editing the xwiki.properties file related to notifications, it does not specifically mention setting 'notifications.enabled' to 'true', which is a key detail in the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications by setting 'notifications.enabled' to 'true' in 'xwiki.properties'.

Correctness: The actual output provides additional information about the XWiki interface and specific extensions but does not address enabling notifications via 'notifications.enabled' in 'xwiki.properties' as stated in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3065569996498207

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.006008664565590819

Reasons:

AnswerRelevancy: The score is 0.83 because the response is mostly relevant to the issue of not receiving notifications, but it includes a source link that does not directly address this specific problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node contains 'a list of image links and references, which are unrelated to the topic of enabling notifications or settings in 'xwiki.properties'.' The second node also consists of 'image links and references, which do not provide any information about notification settings or following users/pages in XWiki.' Similarly, the third node contains 'image links and references, which are not relevant to the instructions on activating notifications in XWiki.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by the nodes in the retrieval context, which lack relevant information about enabling notifications, following pages or users, and related settings.

CustomContextualRelevancy: The retrieval context contains image links and references, but lacks any information about enabling notifications or following pages and users as described in the expected output.

Correctness: The actual output lacks specific instructions on enabling notifications in `xwiki.properties` and following pages or users, which are key facts in the expected output. It instead discusses general notification settings unrelated to the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.21828118736719235

Individual_scores:

AnswerRelevancy: 0.4444444444444444

Faithfulness: 0.6666666666666666

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.19857601309204287

Reasons:

AnswerRelevancy: The score is 0.44 because the response contains multiple irrelevant statements, such as unrelated source links and suggestions, which do not directly address the question about enabling notifications for one's own actions. However, there might be some partial relevance or attempts to address the question, which prevents the score from being lower.

Faithfulness: The score is 0.67 because the actual output incorrectly refers to 'What's New' as an extension, while the retrieval context specifies it as a UI feature called 'What's New UI'.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customization of page templates, internationalization, and other features unrelated to notification settings.' The second node is about 'administrating wiki features and does not mention notification settings or user activity.' The third node covers 'various XWiki extensions like Office Importer, Font Awesome, and others, none of which relate to user notification settings.'

These irrelevant nodes should be ranked lower than any relevant nodes.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information about notification settings or user activity filters mentioned in the expected output.

Correctness: The actual output lacks key facts from the expected output, such as not mentioning the lack of default notifications for one's own activity and how to disable this filter in notification settings. It introduces irrelevant information about features and extensions not present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.2984203734755149

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0
Correctness: 0.2905222408530894
Reasons:

AnswerRelevancy: The score is 0.50 because while the output contains some relevant information about managing notifications, it is diluted by multiple irrelevant links that do not address the specific question about disabling notifications.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files', which does not relate to notification settings. The second node talks about 'moving attachments', which is unrelated to managing notifications. The third node is about 'displaying various elements', not addressing notification settings. The fourth node pertains to 'administrating wiki features', without mentioning notifications. The fifth node focuses on 'controlling permissions', not related to notifications. Lastly, the sixth node lists 'applications and extensions', none of which pertain to notifications.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about automatic page following or notification settings as described in the expected output.

Correctness: The actual output discusses notification settings and page tracking but lacks specific details about automatic tracking of edited pages and deleting custom filters as mentioned in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_015_result.json

Overall_score: 0.5209515616693371
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 0.75
ContextualPrecision: 0.25
ContextualRecall: 0.8
CustomContextualRelevancy: 0.2
Correctness: 0.1257093700160225
Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring XWiki for automatic user activation with OpenID Connect. Great job!

Faithfulness: The score is 0.75 because the actual output incorrectly states that the `oidc.skipped` property must be disabled for OpenID Connect authentication to work, while the retrieval context indicates that OpenID Connect can be disabled using this property.

ContextualPrecision: The score is 0.25 because the relevant node, which mentions the ``xwiki.properties`` file and includes examples of OpenID Connect related properties, is ranked fourth. This node is crucial for configuring automatic user activation. However, it is ranked lower than three irrelevant nodes. The first node discusses domain-based authentication and session cookie configuration, which is unrelated. The second node provides general information about OpenID Connect authentication without addressing the specific configuration needed. The third node contains metadata about the OpenID Connect authenticator, which does not assist in configuring automatic user activation.

ContextualRecall: The score is 0.80 because most sentences in the expected output are supported by the information in the 4th node in the retrieval context, which discusses configuring user accounts with

OpenID Connect in XWiki. However, the instruction to restart the XWiki instance is not explicitly mentioned in any node.

CustomContextualRelevancy: The retrieval context mentions the xwiki.properties file but lacks specific details about enabling user accounts with oidc.enableUser=true.

Correctness: The actual output does not match the expected factual statements. It incorrectly suggests modifying `xwiki.authentication.authclass` and `oidc.skipped` instead of `oidc.enableUser`. The expected change to enable user accounts is not addressed, and additional unrelated suggestions are provided.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.3999316250155333

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.6666666666666666

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.0

Correctness: 0.13292308342653333

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about customizing user profile fields in XWiki using OpenID Connect, with no irrelevant statements present. Great job!

Faithfulness: The score is 0.67 because the actual output incorrectly claims that the OpenID Connect Provider extension is required, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect endpoints', which does not mention user profile field mapping. The second node is about 'container authentication and other methods like LDAP and Active Directory', unrelated to OpenID Connect user profile fields. The third node provides a 'general description of OpenID Connect authentication' without addressing user profile field mapping. The fourth node includes 'metadata about the OpenID Connect Authenticator extension', but does not discuss user profile field mapping. The fifth node lists 'dependencies for the OpenID Connect Authenticator extension', which is unrelated to the customization of user profile fields.

ContextualRecall: The score is 0.60 because while the nodes in retrieval context support the general process of configuring OpenID Connect user mappings and saving changes in 'xwiki.properties', they lack specific examples and instructions on replacing placeholder values, which are present in sentences 2 and 3 of the expected output.

CustomContextualRelevancy: The retrieval context lacks information about customizing user profile fields using `oidc.user.mapping` in `xwiki.properties` as described in the expected output.

Correctness: The actual output fails to provide any of the key facts present in the expected output regarding customization of user profile fields using `xwiki.properties` and `oidc.user.mapping`. It only suggests exploring other resources and does not mention the specific steps or code examples given in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.3918743102193369

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.3

CustomContextualRelevancy: 0.0

Correctness: 0.05124586131602119

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about setting up group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job on maintaining accuracy!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'authentication management and OpenID Connect endpoints' without addressing 'group synchronization or the specific configuration steps needed for it'. The second node talks about 'different authentication methods and extensions' but lacks 'details on group synchronization or configuration in xwiki.properties'. The third node provides links to configuration guides but does not 'detail the group synchronization setup or the specific properties in xwiki.properties'. The fourth node lists 'dependencies for the OpenID Connect extension', which is unrelated to the 'group synchronization configuration process'. Lastly, the fifth node gives 'general information about the OpenID Connect authenticator' but omits 'group synchronization or configuration steps'.

ContextualRecall: The score is 0.30 because while the configurations for 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' are supported by the 1st node in the retrieval context, many procedural instructions like finding and editing the 'xwiki.properties' file, replacing placeholders, and restarting XWiki are not supported by the retrieval context.

CustomContextualRelevancy: The retrieval context lacks specific instructions on configuring group synchronization properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' as detailed in the expected output.

Correctness: The actual output does not align with the expected instructions for setting up group synchronization in XWiki with an OpenID Connect provider. It mentions the lack of direct support and suggests contacting the developer, which is not part of the provided expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.5660588298688176

Individual_scores:

AnswerRelevancy: 0.7647058823529411

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.4444444444444444

CustomContextualRelevancy: 0.5

Correctness: 0.4872026524155202

Reasons:

AnswerRelevancy: The score is 0.76 because the output partially addresses the question about configuring attachment upload restrictions, but includes several irrelevant statements about exports, XWiki syntax, document lifecycle, and default settings, which do not directly relate to the specific

requirements of file type and size limits.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.20 because the relevant node in the retrieval context, which provides 'detailed steps on setting the maximum attachment size and mentions MIME type restrictions', is ranked fifth. Irrelevant nodes, such as the first node discussing 'trusted domains and PDF export customization', and the second node about 'ODT and PDF export features', are ranked higher, which lowers the score.

ContextualRecall: The score is 0.44 because while the 5th node in the retrieval context provides detailed steps for setting maximum upload size, aligning with sentences in the expected output, it lacks information on configuring MIME types and specific size limits for images or PDFs, which are crucial parts of the expected output.

CustomContextualRelevancy: The retrieval context mentions configuring maximum attachment size and MIME type restrictions, but lacks specific details about setting MIME types in xwiki.properties and the exact steps for setting the size to 10MB.

Correctness: The actual output correctly outlines setting the maximum upload size but misses configuring MIME types in xwiki.properties. It mentions altering trusted domains, which isn't in expected output. There's a mention of a default setting of 100 GB that's not in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.33067846619207003

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.26978508286670577

Reasons:

AnswerRelevancy: The score is 0.71 because the response mostly addresses the question about revoking script rights from a space administrator, but includes a source link that does not directly answer the question, which slightly detracts from its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'form validation and JavaScript patterns', the second node is about 'template providers and page creation in XWiki', and the third node involves 'a custom right check and configuration in XWiki', none of which relate to denying script rights to an area administrator.

ContextualRecall: The score is 0.00 because the expected output sentence about admin rights and their limitations in XWiki cannot be linked to any information in the nodes in retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to admin rights or their implications, which is the main focus of the expected answer.

Correctness: The actual output does not align with the expected output. It fails to mention the key factual detail that rights implied by admin rights cannot be denied. Instead, it provides general advice on permission changes and suggests consulting an expert, which deviates from the expected factual statement.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.3128049461966455

Individual_scores:

AnswerRelevancy: 0.5454545454545454

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.125

CustomContextualRelevancy: 0.0

Correctness: 0.20637513172532743

Reasons:

AnswerRelevancy: The score is 0.55 because while the output does address configuring access rights in XWiki, it includes several irrelevant links and information that do not directly relate to the specific task of setting up team-specific access permissions.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'XWiki's database efficiency and programming features,' which are unrelated to configuring access controls for teams. The second node is about 'notifications and application ratings,' which do not pertain to setting up team access permissions in XWiki. The third node covers 'using nested pages and subwikis,' which does not directly address configuring team-specific access rights. The fourth node refers to 'managing users and groups,' but does not provide specific guidance on setting access controls for team areas. Finally, the fifth node discusses 'group synchronization and customization,' which does not directly relate to configuring team access permissions in XWiki.

ContextualRecall: The score is 0.12 because only the concept of creating groups for teams (sentence 1) is supported by the 4th node in the retrieval context, while all other sentences in the expected output lack support from the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to access control steps or rights management as outlined in the expected answer.

Correctness: The actual output fails to provide specific steps for access control configuration, which is a critical element of the expected output. It only vaguely mentions user and group management without detailing steps like creating groups, setting wiki-wide rights, or inheriting rights for sub-pages. It lacks factual accuracy and introduces uncertainty by directing users to external documentation for specific instructions.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.39719142747396835

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.21648189817714383

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the issue of the notification bell not appearing, but includes irrelevant links that do not directly help solve the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'custom notification filters and their behavior', which is not related to the issue of enabling notifications in the wiki settings. Subsequent nodes also fail to address the relevant topic, such as the second node listing 'various notification-related features and links' without mentioning the enabling of notifications in the wiki settings.

ContextualRecall: The score is 0.50 because the retrieval context supports the need for activating notifications as mentioned in sentence 1, but lacks information about setting 'notifications.enabled' in 'xwiki.properties' as described in sentence 2.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the 'xwiki.properties' file or setting 'notifications.enabled' to 'true'.

Correctness: The actual output does not mention enabling notifications in 'xwiki.properties' as expected. It introduces new information about the 'Application de notifications' and its icon, which is not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.7014513485191968

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.875

CustomContextualRelevancy: 0.4

Correctness: 0.21942237682946616

Reasons:

AnswerRelevancy: The score is 0.71 because the output partially addresses the issue of not receiving notifications, but includes irrelevant source links that do not directly relate to the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes are perfectly ranked above the irrelevant ones. Great job!

ContextualRecall: The score is 0.88 because most of the expected output is supported by nodes in the retrieval context, such as enabling notifications and following pages or users, which are covered by the 1st and 2nd nodes in the retrieval context. However, there is no mention of 'xwiki.properties' or setting 'notifications.enabled', which slightly affects the score.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' to 'true' in 'xwiki.properties' and specific steps for following pages or users.

Correctness: The actual output fails to mention enabling notifications in 'xwiki.properties' and lacks guidance on following pages or users. It includes unrelated information about default settings and notification types, which are not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8774531633115455

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Correctness: 0.586941202091495

Reasons:

AnswerRelevancy: The score is 0.78 because the response mostly addresses the question about enabling notifications for one's own actions, but includes irrelevant source links that do not directly contribute to solving the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, ranked first, directly addresses the input question by stating, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter.' All irrelevant nodes are ranked lower, discussing unrelated topics such as 'getting a notification RSS feed' (second node), 'a list of links' (third node), 'various notification preferences' (fourth node), and 'various settings and features' (fifth node), none of which address the input question.

ContextualRecall: The score is 1.00 because all sentences in the expected output are fully supported by nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context confirms that users do not receive notifications for their own actions by default and mentions the ability to change this setting, aligning closely with the expected output. However, the specific location 'Options de filtrage avancées' is not mentioned in the retrieval context.

Correctness: The actual output provides the key fact that users can disable filtering to receive notifications about their activities, aligning with the expected output. However, it lacks the detail about the default setting of not receiving such notifications unless explicitly targeted. Steps for changing settings are mentioned but not the specific section 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7748094256850534

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Correctness: 0.24171369696746323

Reasons:

AnswerRelevancy: The score is 0.86 because the response mostly addresses the question about disabling notifications, but includes irrelevant links that do not pertain to the query, slightly lowering the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all nodes in the retrieval context are relevant and appropriately ranked, addressing the input query about disabling automatic notifications with precise information on notification settings and preferences.

ContextualRecall: The score is 0.75 because nodes in the retrieval context support the expected output's details about automatic page tracking and modifying notification settings (sentences 1, 3, and 4). However, there is no mention in the retrieval context about being notified of changes to your work (sentence 2).

CustomContextualRelevancy: The retrieval context includes information about auto-watched pages and custom notification filters, which aligns with the expected answer. However, it lacks explicit details about modifying notification settings in the user profile.

Correctness: Actual output lacks emphasis on automatic tracking of pages after significant edits and does not mention altering 'Suivi automatique des pages' specifically. It introduces unrelated information about notification types and global filters instead.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.3890087665223921

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.05905259913435257

Reasons:

AnswerRelevancy: The score is 0.88 because the response is mostly relevant to configuring OpenID Connect in XWiki, but it includes an irrelevant mention of LDAP Authenticator, which is not related to the question about OpenID Connect.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval contexts are irrelevant to the input. The first node discusses 'domain-based authentication and session cookies', which is unrelated to enabling user accounts via OpenID Connect in XWiki. The second node is about 'container authentication and LDAP', not addressing the configuration of OpenID Connect for automatic user activation. The third node mentions OpenID Connect but lacks information on enabling user accounts automatically upon first login. The fourth node lists 'dependencies for the OpenID Connect extension', which is unrelated to the specific configuration needed. Lastly, the fifth node focuses on 'LDAP and Active Directory integration', which is not relevant to the OpenID Connect configuration for automatic user activation.

ContextualRecall: The score is 0.40 because while nodes in the retrieval context mention 'OpenId Connect Authentication' and 'xwiki.properties', they do not specifically address adding or updating the property for enabling user accounts, restarting XWiki, or automatically activating accounts on first login, which are key parts of the expected output.

CustomContextualRelevancy: The retrieval context lacks any mention of the xwiki.properties file or the oidc.enableUser property, which are crucial details in the expected answer.

Correctness: The actual output does not match the expected output; it lacks the correct property `oidc.enableUser=true` for activating accounts and suggests incorrect configurations and use of extensions or scripts, which are not mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.45801046955594593

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.4147294840023422

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to the question about mapping additional user profile fields from an OpenID Connect provider to XWiki user properties, with only a minor mention of automatic synchronization that slightly detracts from the focus.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management class and OpenID Connect endpoints', the second node is about 'different authentication methods', the third node describes 'OpenID Connect Authenticator extension', the fourth node lists 'dependencies for the OpenID Connect Authenticator extension', and the fifth node provides 'configuration guides and limitations for OpenID Connect'. None of these nodes address the mapping or customization of user profile fields, which is the main focus of the input.

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention 'xwiki.properties' and OpenID Connect, which supports the customization of user profile fields, they lack specific examples of mapping properties, instructions on replacing property names or variables, and do not discuss saving the file or restarting XWiki.

CustomContextualRelevancy: The retrieval context does not contain information about customizing user profile fields using `oidc.user.mapping` in `xwiki.properties` as mentioned in the expected output.

Correctness: The actual output acknowledges the possibility of customizing user profile fields with OpenID Connect but lacks specific mapping instructions using `oidc.user.mapping` as in the expected output. It also mentions contacting a developer, which is not in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5437037486302044

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.3076923076923077

CustomContextualRelevancy: 0.3

Correctness: 0.32119685075558513

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider.

Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, ranked third, provides information about group synchronization, mentioning the use of claims and the `oidc.groups.mapping` property. However, irrelevant nodes are ranked higher, such as the first node which discusses authentication management and endpoint configurations without mentioning group synchronization, and the second node which talks about different authentication methods without providing information on configuring group synchronization with OpenID Connect.

ContextualRecall: The score is 0.31 because while the 3rd node in the retrieval context supports the general concept of group synchronization and configuration in XWiki, it lacks specific details such as locating and editing the 'xwiki.properties' file, providing examples, and saving changes, which are crucial for a complete match with the expected output.

CustomContextualRelevancy: The retrieval context mentions group synchronization and oidc.groups.mapping but lacks details on configuring oidc.groups.claim and oidc.groups.separator, which are present in the expected output.

Correctness: Actual output lacks specific steps such as configuring `oidc.groups.claim` and `oidc.groups.separator`. It introduces `oidc.userinfoclaims` which is not in the expected output and adds a custom synchronization step which is not required. However, it correctly mentions mapping groups and restarting XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.2776632282968908

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.16597936978134467

Reasons:

AnswerRelevancy: The score is 0.50 because the output contains multiple irrelevant source links that do not directly address the question about limiting file uploads in XWiki, which reduces the relevancy. However, the score is not lower because there is still some relevant information addressing the input question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'language preferences and translations in XWiki', which is unrelated to file upload restrictions. The second node talks about 'various administrative applications and extensions in XWiki', not addressing file upload settings. The third node focuses on 'PDF export and styling options in XWiki', which does not relate to configuring file upload restrictions. The fourth node lists 'various system preferences and macros in XWiki', none of which pertain to file upload restrictions. The fifth node contains 'links to image files and references', which are not relevant to the topic of configuring file upload restrictions. The sixth node also provides 'links to image files and references', unrelated to the configuration of file upload restrictions in XWiki.

ContextualRecall: The score is 0.00 because none of the expected output sentences can be attributed to any node(s) in the retrieval context, as the context lacks information on file upload restrictions, MIME

type configurations, and attachment size settings.

CustomContextualRelevancy: The retrieval context does not contain any information related to configuring MIME types or attachment size settings in XWiki as described in the expected output.

Correctness: The actual output does not match the expected output; it suggests using a PHP configuration change, which is not mentioned in the expected steps. It also incorrectly states no built-in method exists for limiting attachment size in XWiki when the expected output provides a method.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.6010725806562659

Individual_scores:

AnswerRelevancy: 0.7

Faithfulness: 0.8333333333333334

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.2897688172709285

Reasons:

AnswerRelevancy: The score is 0.70 because the response addresses the main question about refusing script rights to a space administrator, but it includes irrelevant source links that do not contribute to answering the specific query.

Faithfulness: The score is 0.83 because the actual output incorrectly claims that the script right is granted by default to all users with edit rights, whereas the retrieval context clarifies that the script right is allowed for all users at the main wiki level, not specifically linked to edit rights.

ContextualPrecision: The score is 0.58 because relevant nodes discussing programming and script rights (second node) and the 'ADMIN' right being 'not deniable' (third node) are ranked after an irrelevant node about language preferences and translations in XWiki (first node). Additionally, another irrelevant node about checking access rights programmatically (fourth node) is ranked higher than the relevant node about the 'ADMIN' right. However, the relevant nodes are still ranked higher than some other irrelevant nodes, which contributes to the score being above 0.5.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the nodes in the retrieval context, specifically the 2nd node, which aligns perfectly with the discussed topic.

CustomContextualRelevancy: The retrieval context mentions rights like ADMIN and programming that cannot be denied, but lacks explicit mention of XWiki not supporting refusal of admin rights.

Correctness: The actual output acknowledges that refusing script rights to a space admin is not typically supported, aligning partially with the expected output. However, it introduces incorrect info about controlling script rights via the security API, which is inconsistent with the expected output stating it's not supported.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.3028555747150284

Individual_scores:

AnswerRelevancy: 0.6

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2171334482901704

Reasons:

AnswerRelevancy: The score is 0.60 because while the output provides source links, it fails to directly address the specific question about configuring spaces and permissions in XWiki, leading to repeated irrelevant statements.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining accuracy and alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node describes an XML page structure, which is unrelated to configuring access rights in XWiki. The second node discusses sending representations and modifying resources using XML and HTTP methods, which does not pertain to setting up access rights for spaces in XWiki. The third node is about editing comments and managing attachments, which is not related to configuring access rights for team spaces in XWiki. The fourth node explains spaces and sub-pages in XWiki, but does not provide guidance on setting access rights for different teams. The fifth node provides information about migrating and updating XWiki configurations, but does not address setting access rights for team spaces. Finally, the sixth node outlines configuration display options in XWiki, but does not relate to setting access rights for spaces as required by the input question.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, defining rights at the wiki or space level, or inheriting rights for child pages as mentioned in the expected output.

Correctness: The actual output does not address creating groups or setting rights at the wiki or space level, and instead mentions page-level permissions. It fails to mention inheriting rights for child pages, leading to incomplete and misaligned information.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.6891510956721144

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.7

Correctness: 0.5063351454612569

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about authenticating users with access tokens, but includes multiple irrelevant statements providing source links that do not directly answer the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because relevant nodes are not consistently ranked higher than irrelevant nodes. The first node discusses 'LDAP authentication settings, which is unrelated to token-based authentication with JWTs' and should be ranked lower. The second node is relevant as it 'provides information on configuring authorized applications, which is relevant to setting up token-based authentication.' The third node, which 'describes OpenID Connect scopes and response types, which

do not directly relate to JWT token-based authentication for the LLM Application,' should also be ranked lower. The fourth node is relevant as it 'describes the JWT token authenticator, which is directly relevant to the expected output about using access tokens for authentication.' Finally, the fifth node 'provides a general overview of XWiki authentication mechanisms, but does not specifically address JWT token-based authentication,' and should be ranked lower.

ContextualRecall: The score is 0.71 because while the retrieval context nodes provide substantial information about JWT authentication, such as enabling the token authenticator and configuring authorized applications, they lack specific details on generating a JWT token with required claims and including it in the 'Authorization' header.

CustomContextualRelevancy: The retrieval context mentions the use of JWT for authentication and configuring authorized applications, aligning with the expected answer. However, it lacks details on specific claims required in the token and the use of Ed25519 key for signing.

Correctness: The actual output mentions JWT and how to enable it, aligning partly with the expected output. However, it introduces OpenID Connect, which is additional incorrect information not present in the expected output. It also lacks details on configuring authorized applications, generating JWT tokens with specific claims, and the limitation regarding user creation.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.2954501443525957

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.022700866115574363

Reasons:

AnswerRelevancy: The score is 0.75 because the main content addresses the question about authenticating users with access tokens, but the inclusion of irrelevant source links prevents a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'model validation strategies and configuration in XWiki', the second node is about 'attachment validation and configuration in XWiki', and the third node provides information on 'the admin UI for model validation'. None of these are related to 'authenticating users with access tokens', which is the focus of the input.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output have any corresponding information in the nodes in retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which is central to the expected output.

Correctness: The actual output does not provide any of the specific instructions or details found in the expected output about token-based authentication using JWT in XWiki. It instead discusses general authentication methods like OpenID or LDAP, which are not relevant.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.30882192487915033

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.102931549274902

Reasons:

AnswerRelevancy: The score is 0.75 because the response mostly addresses the question about authenticating users with access tokens, but includes irrelevant source links that do not directly relate to the topic.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'upgrading XWiki' without mentioning 'authentication or access tokens'. The second node talks about 'features of the Administration Application', focusing on 'user rights and registration' without referencing 'token-based authentication'. The third node is about 'managing user rights and profiles', but it does not address 'authenticating users with access tokens'. Thus, none of the nodes provide relevant information for the input query.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context. The expected output discusses token-based authentication and JWTs, which are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are central to the expected answer.

Correctness: The actual output does not mention the token-based authentication extension or JWT requirements, which are key facts in the expected output. It talks about general authentication and rights management, which is unrelated to the specific criteria.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_gemma2_9B_Q4

File: qa_001_result.json

Overall_score: 0.3153455734980692

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.009720499811944549

Reasons:

AnswerRelevancy: The score is 0.88 because the response is mostly relevant and helpful, addressing the issue of the missing 'Bell' notification icon. However, it includes repeated irrelevant statements about a source link that do not directly contribute to solving the problem, preventing a perfect score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'installation instructions and prerequisites for extensions,' which does not address enabling notifications in XWiki. The second node describes 'the functionality of the notifications and the bell icon' but lacks information on enabling them. The third node lists dependencies unrelated to enabling notifications. Lastly, the fourth node covers 'overriding notification templates,' which is not helpful for the issue at hand.

ContextualRecall: The score is 0.00 because the retrieval context does not provide any information related to enabling notifications in the wiki or the specific settings mentioned in the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the `notifications.enabled` setting in `xwiki.properties`.

Correctness: The actual output does not address enabling notifications via the `notifications.enabled` setting in `xwiki.properties` as described in the expected output. It discusses troubleshooting the Notifications Application and potential issues, which is unrelated to the expected task.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6861828360491643

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.6

Correctness: 0.26709701629498606

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining accuracy and alignment with the retrieval context!

ContextualPrecision: The score is 0.75 because the first node in the retrieval context is relevant, explaining how to follow a user to receive notifications, and the fourth node is also relevant, providing an overview of ways to receive notifications in XWiki. However, the second node, discussing filters and settings without addressing enabling notifications, and the third node, focusing on RSS feeds and settings, are both irrelevant and ranked higher than the fourth relevant node. Additionally, the fifth node, which is about the 'What's New' application, is not directly related to enabling notifications.

ContextualRecall: The score is 0.50 because half of the sentences in the expected output are well-supported by nodes in the retrieval context, such as the details on using the 'alert' menu and following a user. However, key setup instructions like enabling notifications in `xwiki.properties` and subscribing to pages are missing from the retrieval context.

CustomContextualRelevancy: The retrieval context includes information on following users and watching pages, but lacks details on enabling notifications in `xwiki.properties` and watching entire wikis.

Correctness: The actual output mentions enabling notification types and checking filters, which partially aligns with enabling notifications in the expected output. However, it lacks details about setting notifications in `xwiki.properties`, watching pages or users, and using the 'alert' menu. No mention of the 'network' tab in user profiles is made.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.8575288779709581

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 0.6666666666666666

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.8118399344924155

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses how to enable notifications for one's own actions in XWiki, but includes irrelevant source links that do not directly contribute to solving the issue.

Faithfulness: The score is 0.67 because the actual output incorrectly claims that the 'Own event filter' is usually enabled by default, while the retrieval context indicates that users do not receive notifications for their own actions by default, suggesting that the filter is indeed enabled by default.

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than irrelevant nodes. The first and second nodes provide precise information about enabling notifications for one's own actions in XWiki, while the third, fourth, and fifth nodes, which focus on administrative settings, RSS feeds, and general notifications, are correctly ranked lower as they do not address the specific issue.

ContextualRecall: The score is 1.00 because all sentences in the expected output are fully supported by the nodes in the retrieval context, showcasing a perfect alignment. Great job!

CustomContextualRelevancy: The retrieval context fully contains the expected answer, mentioning that by default, notifications about one's own activity are not received unless explicitly targeted, and this can be changed in the 'Advanced filtering options' of the user's notification settings.

Correctness: Both outputs mention that by default, one does not receive notifications about their own activity and that this can be changed in the notification settings. However, the actual output includes additional information about the specific filter name 'Own event filter' not mentioned in the expected output, but this does not contradict the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.608561647961182

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Correctness: 0.351369887767092

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about disabling notifications in XWiki without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'getting a notification RSS feed and enabling/disabling notification types', but it does not address 'how to disable notifications for pages you are not interested in'. The

second node talks about 'following a user and default watch settings', which does not provide information on 'how to disable notifications for unwanted pages'. Subsequent nodes similarly fail to address the specific need of disabling notifications for unwanted pages, focusing instead on unrelated topics such as 'filtering notifications for hidden pages', 'enabling/disabling email notifications', 'administrative settings for notifications', and 'overriding default notification templates'.

ContextualRecall: The score is 1.00 because all sentences in the expected output are fully supported by the nodes in the retrieval context, demonstrating a perfect match. Great job!

CustomContextualRelevancy: The retrieval context mentions default notification settings and user profile settings but lacks details on major modifications automatically adding pages to watchlist and deleting custom filters.

Correctness: The actual output provides a general guide to notification settings in XWiki but lacks specific mention of 'Automatic page watching' and 'Custom Filters' as stated in the expected output. It does not address the automatic addition of major modification pages to the watchlist or the deletion of filters created that way.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8866038555018895

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.9

Correctness: 0.896546209934414

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant to configuring OpenID Connect in XWiki, with only a minor mention of LDAP authentication that slightly detracts from the focus.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job on maintaining accuracy!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions '# oidc.enableUser=true', is ranked first. This directly relates to enabling user accounts on first login, as required in the input. The irrelevant nodes, discussing topics like domain-based authentication, general OpenID Connect information, container authentication, and LDAP configuration, are ranked lower, ensuring the most pertinent information is prioritized.

ContextualRecall: The score is 0.60 because while some sentences in the expected output are supported by the 1st node in the retrieval context, others lack direct support from any node, leading to a moderate recall score.

CustomContextualRelevancy: The retrieval context includes the commented property '# oidc.enableUser=true', matching the expected answer's requirement to uncomment and set it to true, but lacks explicit instructions to save and restart XWiki.

Correctness: The actual output aligns with the expected output, detailing the modification of the 'oidc.enableUser' property and restarting XWiki. However, it does not mention uncommenting the property.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.8491670649988531

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Correctness: 0.7450023899931193

Reasons:

AnswerRelevancy: The score is 0.75 because the answer partially addresses the question about customizing user profile fields in XWiki, but includes irrelevant statements like a source link and a reference to documentation that do not directly answer the question.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user' and provides an example of mapping, is ranked first. This directly relates to customizing user profile fields, while the irrelevant nodes, such as those discussing 'scopes and response types for OpenID Connect' (second node), 'endpoints and HTTP headers for OpenID Connect' (third node), 'group synchronization and customization templates' (fourth node), and 'in-wiki configuration and multiple configurations through cookies' (fifth node), are ranked lower. Great job on getting the order spot on!

ContextualRecall: The score is 0.60 because while the first node in the retrieval context supports the customization of user profile fields and provides examples of mapping them, it lacks details on replacing placeholders with actual property names and the need to save and restart XWiki.

CustomContextualRelevancy: The retrieval context includes the information about customizing user profile fields using the `oidc.user.mapping` property in the `xwiki.properties` file, which matches the expected output.

Correctness: Both outputs describe using the `oidc.user.mapping` property in `xwiki.properties` for customizing user profile fields from OpenID Connect. However, the actual output uses a different example and omits the restart step mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8252876045024932

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Correctness: 0.5517256270149582

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider, without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant ones. The first three nodes provide detailed information on configuring the `oidc.groups.claim` and `oidc.groups.mapping` properties, which are crucial for group synchronization as required by the input. The irrelevant nodes, ranked fourth and fifth, focus on general configuration and authentication management, which do not specifically address the group synchronization process.

ContextualRecall: The score is 0.50 because while there is a good match for configuring properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` with nodes in the retrieval context, key steps such as locating and opening the `xwiki.properties` file, replacing placeholder values, and saving the file are not supported by the retrieval context.

CustomContextualRelevancy: The retrieval context covers key aspects of group synchronization such as setting `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, but lacks explicit instructions for locating and editing the `xwiki.properties` file.

Correctness: The actual output mentions configuring group synchronization and restarting XWiki, aligning partially with the expected output. However, it lacks details on setting the `oidc.groups.claim` and `oidc.groups.separator` properties, and uses `oidc.userinfoclaims` instead, which isn't in the expected output.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_008_result.json

Overall_score: 0.6387772333636667
Individual_scores:
AnswerRelevancy: 0.9473684210526315
Faithfulness: 1.0
ContextualPrecision: 0.3333333333333333
ContextualRecall: 0.8
CustomContextualRelevancy: 0.2
Correctness: 0.5519616457960351

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant and addresses the question about configuring attachments, but it includes a link to general information that might not directly answer the specific query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which provides information on setting the maximum size of an attachment and mentions the configuration parameter in the XWikiPreferences document, is ranked third. However, irrelevant nodes, such as the first node discussing linking to external files and images, and the second node describing the structure of XWiki pages, are ranked higher. These nodes do not address configuring upload restrictions or size limits, affecting the precision score.

ContextualRecall: The score is 0.80 because most of the detailed steps in the expected output are well-supported by nodes in the retrieval context, particularly nodes 3 and 4, which cover configuring mimetypes and setting maximum attachment sizes. However, the introductory and explanatory sentences lack direct support from specific nodes in the retrieval context.

CustomContextualRelevancy: Retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific details on setting size to 10MB and configuring mimetypes in `xwiki.properties`.

Correctness: The actual output correctly sets the maximum upload size to 10MB and uses a whitelist for allowed mimetypes, but it specifies explicit image types instead of `image/*`, potentially missing some subtypes.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_009_result.json

Overall_score: 0.6510085734578082

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.10605144074684911

Reasons:

AnswerRelevancy: The score is 0.80 because the main content effectively addresses how to deny script rights to a space administrator, but the inclusion of multiple source links that do not directly answer the question slightly detracts from the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions. Everything aligns perfectly with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses programming rights and their implications, but does not address the denial of script rights to a space administrator. The second node provides a list of rights and their properties, but does not mention the inability to deny script rights to a space administrator. The third node explains the script right and its default settings, but does not mention the inability to deny it to a space administrator. The fourth node is about checking access rights programmatically, which does not relate to denying script rights to a space administrator. The fifth node discusses converting rights during page migration, which is unrelated to denying script rights to a space administrator.

ContextualRecall: The score is 1.00 because every part of the expected output is fully supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The expected answer is fully supported by the retrieval context, which states that admin rights imply other rights and cannot be denied, aligning with the expected output.

Correctness: The actual output provides a method to deny script rights to a space administrator, contradicting the expected output that states rights implied by admin cannot be denied.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_010_result.json

Overall_score: 0.753344863004887

Individual_scores:

AnswerRelevancy: 0.92

Faithfulness: 0.875

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.875

CustomContextualRelevancy: 0.6

Correctness: 0.5834025113626561

Reasons:

AnswerRelevancy: The score is 0.92 because the response is mostly relevant and addresses the question about configuring permissions in XWiki. However, it includes irrelevant source links that do not directly relate to the configuration of permissions, which slightly lowers the score.

Faithfulness: The score is 0.88 because the actual output incorrectly claims that XWiki uses a hierarchical permission system where higher-level permissions are inherited by lower levels, contradicting the retrieval context which states that page-level permissions can override wiki-wide permissions.

ContextualPrecision: The score is 0.67 because the first node in the retrieval context is relevant, discussing 'rights inheritance and administration features' and 'Inheritance of access rights', which are pertinent to configuring view and edit permissions in XWiki. However, the second node, ranked higher than some relevant nodes, is irrelevant as it focuses on the 'Nested Pages concept' and naming conflicts, not on permissions. Similarly, the third node is unrelated, addressing string formatting and user identifiers, which do not pertain to access control. The fourth node is relevant, detailing XWiki's permission system with 'View' and 'Edit' rights, but it is ranked lower than the irrelevant nodes. Lastly, the sixth node is relevant, explaining 'wiki wide rights, granular page level rights', and permission priorities, but it is also ranked lower than some irrelevant nodes.

ContextualRecall: The score is 0.88 because most of the expected output aligns well with the nodes in the retrieval context, such as creating groups, setting wiki-wide and space-level rights, and inheriting rights for child pages. However, there is no specific mention in the nodes about adjusting rights for additional teams or spaces, which slightly affects the score.

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and inheriting rights for child pages, aligning with steps 2 and 4 of the expected output. However, it lacks specific instructions on creating groups for each team and setting space-level rights, which are crucial parts of the expected answer.

Correctness: The actual output covers group creation and space-level permissions, aligning with expected steps 1 and 3. However, it lacks explicit mention of wiki-wide rights and inheritance for child pages, and includes additional details on explicit denials and an admin guide not present in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.4207677970758554

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.1079401157884653

Reasons:

AnswerRelevancy: The score is 0.75 because the response addresses the notification bell issue, but includes transitional and closing remarks that do not directly contribute to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node, which mentions editing the `xwiki.properties` file to enable notifications, is ranked sixth. The first five nodes are irrelevant as they discuss topics like 'custom notification filters', 'alert menu', 'ways to receive notifications', 'different notification settings', and 'customizing notification settings', none of which address enabling notifications in the settings. This lowers the score since the relevant information is not prioritized.

ContextualRecall: The score is 0.50 because while the retrieval context (node 6) supports the mention of editing the xwiki.properties file, it does not specifically mention setting 'notifications.enabled' to 'true', which is crucial for the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via `notifications.enabled` in `xwiki.properties`, which is the core information in the expected answer.
Correctness: The actual output does not address enabling notifications via `xwiki.properties`, and instead focuses on installation of the Notifications Application, which is not mentioned in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: en

File: qa_012_result.json

Overall_score: 0.16666666666666666

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because the output contains irrelevant information, such as a link to XWiki skins documentation and a request for the user's question, which do not address the issue of not receiving notifications.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant, containing 'image links and references that are not related to notifications or their settings' (first node), 'do not provide any information about notification settings or how to enable them' (second node), and 'include image links and references that are unrelated to the topic of notifications or their configuration' (third node).

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context contains no relevant information about enabling notifications or following pages and users as described in the expected output.

Correctness: The actual output does not contain any relevant factual statements or information from the expected output. It is entirely missing the instructions on enabling notifications and following pages or users.

Question_language: de
Expected_answer_language: de
Answer_language: en

File: qa_013_result.json

Overall_score: 0.22540403169507142

Individual_scores:

AnswerRelevancy: 0.3333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.01909085683709543

Reasons:

AnswerRelevancy: The score is 0.33 because the output includes irrelevant source links that do not directly address the question about activating notifications for one's own activity.

Faithfulness: The score is 1.00 because there are no contradictions, indicating the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customization of page templates, internationalization, and other features unrelated to notification settings', the second node is about 'administrating wiki features and does not mention notification settings or user activity', and the third node covers 'various XWiki extensions like Office Importer and Syntax Highlighting, none of which relate to notifications or user activity settings'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user profile options, which are central to the expected output.

Correctness: The actual output does not contain any of the expected factual information about notification settings or how to modify them, and instead, it provides unrelated source links and an apology, indicating a complete mismatch.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_014_result.json

Overall_score: 0.28833175054972043

Individual_scores:

AnswerRelevancy: 0.5454545454545454

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.18453595784377735

Reasons:

AnswerRelevancy: The score is 0.55 because while the output contains some relevant information about disabling notifications, it is diluted by multiple irrelevant links to documentation pages and extensions that do not directly address the query about disabling notifications.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are about topics such as 'attaching files and managing attachments', 'moving and deleting attachments', 'macros and functionalities', 'administrating wiki features', 'permissions and security modules', and 'XWiki extensions', which are unrelated to the input about disabling notifications. Therefore, irrelevant nodes are ranked higher than relevant ones.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about automatic page following, notification settings, or notification filters as mentioned in the expected output.

Correctness: The actual output suggests adjusting notification settings in general but misses the specific fact about auto-following pages with significant changes and managing custom filters as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_015_result.json

Overall_score: 0.39255997550143595

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Correctness: 0.23869318634194897

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant and addresses the main question about configuring XWiki for automatic user activation with OpenID Connect. However, it includes a source link that doesn't directly solve the configuration issue, slightly affecting the score.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based authentication and session cookie configuration,' which is unrelated to the automatic activation of user accounts upon first login with OpenID Connect. The second node provides 'general information about OpenID Connect authentication and configuration guides,' but does not address the specific configuration needed for automatic activation. The third node contains 'metadata about the OpenID Connect Authenticator extension,' which does not relate to the configuration of automatic user activation. Lastly, the fourth node provides information about 'various OpenID Connect endpoints and configurations,' but does not mention the specific property 'oidc.enableUser' or its configuration in 'xwiki.properties.'

ContextualRecall: The score is 0.20 because the retrieval context only partially supports the expected output by mentioning modifying properties in the xwiki.properties file (node 4), but it lacks specific details about automatically activating user accounts with OpenID Connect and the exact property changes required.

CustomContextualRelevancy: The retrieval context lacks information on adjusting 'xwiki.properties' to enable automatic user account activation with OpenID Connect.

Correctness: The actual output lacks the key instruction to uncomment and set 'oidc.enableUser=true' in 'xwiki.properties', and instead adds unrelated configuration details. It omits the step to restart XWiki and the assurance that accounts will be activated automatically.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.5842987425374996

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.5057924552249972

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addressed the question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node about the 'xwiki.properties' file, which is crucial for configuring OpenID Connect in XWiki, is ranked first. The irrelevant nodes, such as those discussing 'container authentication and LDAP' (second node), 'general information about OpenID Connect authentication' (third node), 'general information about the OpenID Connect Authenticator extension' (fourth node), and 'dependencies for the OpenID Connect Authenticator extension' (fifth node), are all ranked lower, ensuring the most pertinent information is prioritized. Great job on the ranking!

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any node(s) in the retrieval context, as all key instructions and examples are missing.

CustomContextualRelevancy: The retrieval context does not contain any information about customizing user profile fields using 'xwiki.properties' or 'oidc.user.mapping', which are key aspects of the expected output.

Correctness: The actual output describes customizing user profile fields from OpenID Connect in XWiki, matching the expected output's intent. However, it incorrectly references 'oidc-authenticator-user' instead of 'xwiki.properties' and omits specific variable examples like 'myxproperty1'.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.3858099883732522

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.31485993023951336

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about setting up group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'general OpenID Connect provider configuration and endpoints', but lacks details on group synchronization. The second node is about 'container authentication and other authentication methods like LDAP', which does not relate to the required group synchronization with OpenID Connect. The third node provides 'links and descriptions for OpenID Connect authentication', but misses specifics on configuring group synchronization. The fourth node lists 'dependencies for the OpenID Connect authenticator extension', which is unrelated to group synchronization configuration. Lastly, the fifth node talks about 'installation and compatibility' of the OpenID Connect authenticator extension, without addressing group synchronization setup.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, indicating a complete lack of relevant information.

CustomContextualRelevancy: The retrieval context lacks specific details about configuring group synchronization properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' in 'xwiki.properties'.

Correctness: The actual output mentions configuring the OpenID Connect Authenticator and testing synchronization, but lacks specific file details and properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' as outlined in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.5522328570761704

Individual_scores:

AnswerRelevancy: 0.7647058823529411

Faithfulness: 1.0

ContextualPrecision: 0.325

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Correctness: 0.523691260104081

Reasons:

AnswerRelevancy: The score is 0.76 because the response partially addresses the input question about configuring upload restrictions by providing relevant information. However, the presence of irrelevant source links that do not pertain to the specific question about upload restrictions prevents the score from being higher.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.33 because the first three nodes in the retrieval context are irrelevant, discussing topics like 'trusted domains and PDF export customization', 'ODT export and PDF export applications', and 'various XWiki features and macros', none of which relate to configuring attachment upload restrictions. The relevant nodes, which mention 'Size Limit' and 'Mimetype Restriction' and provide detailed steps on setting these restrictions, are ranked fourth and fifth. Ideally, these relevant nodes should be ranked higher than the irrelevant ones to improve the score.

ContextualRecall: The score is 0.50 because while the retrieval context's 5th node provides detailed steps for configuring the maximum attachment size, it lacks information on MIME type configuration and the specific 10 MB size limit for uploads.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and MIME type restrictions, but lacks specific details on configuring xwiki.properties and setting size to 10 MB as in the expected output.

Correctness: The actual output partially aligns with the expected output but omits specific configuration steps like editing xwiki.properties for MIME types and does not include exact byte value change steps. It introduces an additional method using the Attachment Validation Application not present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.3065711368779323

Individual_scores:

AnswerRelevancy: 0.6

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0
Correctness: 0.23942682126759385

Reasons:

AnswerRelevancy: The score is 0.60 because the output includes a source link, which is helpful, but it fails to directly address the specific question of how to deny script rights to a range administrator.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'form validation and JavaScript patterns', which are unrelated to script rights in XWiki. The second node is about 'templates and page creation in XWiki', not addressing script rights or administrative permissions. The third node involves 'a custom right check and configuration in XWiki', but it does not specifically address denying script rights to a space administrator.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention anything about 'Rechte' or 'Admin-Recht', making it impossible to attribute the expected output to the context.

CustomContextualRelevancy: The retrieval context does not contain any information related to the admin rights or their implications in XWiki, as mentioned in the expected output.

Correctness: The actual output does not address the specific information in the expected output about admin rights not being deniable in XWiki. It instead provides unrelated sources and suggests consulting documentation, missing the key factual point.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_020_result.json

Overall_score: 0.40439949367002986

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.125

CustomContextualRelevancy: 0.0

Correctness: 0.6347302953535126

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the question about configuring access control in XWiki, but it includes multiple irrelevant source links, which do not contribute to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions. Everything aligns perfectly with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'XWiki's database support and programming features', which do not pertain to configuring access control for teams. The second node is about 'notifications and application details in XWiki', unrelated to team access permissions. The third node explains 'nested pages and subwikis', but not access control. The fourth node mentions 'managing users and groups', yet lacks specifics on team access rights. Lastly, the fifth node covers 'group synchronization and customization', without addressing team access permissions.

ContextualRecall: The score is 0.12 because only sentence 1 in the expected output can be linked to the 4th node in the retrieval context, which mentions managing users and groups. All other sentences lack corresponding information in the nodes in retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information related to access control steps or rights management mentioned in the expected output.

Correctness: Actual output aligns well with expected output on creating groups and setting edit rights for team areas. However, it misses details on setting wiki-wide view rights for XWikiAllGroup and inheriting rights for subpages. Also, it introduces additional steps like assigning users to groups and creating pages not explicitly mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_021_result.json

Overall_score: 0.2726850498532594

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 0.8

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.036110299119556565

Reasons:

AnswerRelevancy: The score is 0.80 because the response addresses the issue of the bell icon not being available, but includes irrelevant links and a general offer for assistance that do not directly resolve the specific problem.

Faithfulness: The score is 0.80 because the actual output incorrectly claims that the bell icon is not available, while the retrieval context clearly states that the Notifications Application contributes content to the Alert menu, represented by a bell.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'custom notification filters and their behavior', which is unrelated to 'enabling notifications in the wiki settings'. The second node lists 'various notification settings and preferences' but does not mention 'enabling notifications via xwiki.properties'. The third node describes 'different notification preferences and macros' without addressing 'enabling notifications in the wiki settings'. The fourth node explains 'the notification menu and its features' but fails to mention 'enabling notifications in the wiki settings'. Lastly, the fifth node discusses 'RSS feeds and notification settings' without providing information about 'enabling notifications in the wiki settings'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the 'xwiki.properties' file or setting 'notifications.enabled' to 'true', which is the core information in the expected answer.

Correctness: The actual output provides information about the Notifications Application and potential configuration issues, which are not mentioned or required in the expected output. The expected output solely addresses enabling notifications via xwiki.properties, which is absent in the actual output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_022_result.json

Overall_score: 0.6177555897747039

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Correctness: 0.12875576087044624

Reasons:

AnswerRelevancy: The score is 0.78 because the response mostly addresses the issue of not receiving notifications, but includes irrelevant links that do not directly help resolve the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes are perfectly ranked higher than irrelevant nodes. The first node discusses 'Enable and disable the extension' and 'Enabling notifications for all users by default', which directly relates to ensuring notifications are activated. The second node includes 'Follow a user', aligning with the need to follow users for notifications. Irrelevant nodes, such as the third node discussing 'Get a notification RSS feed' and 'Notifications Macro', and the fourth node listing various notification preferences macros, are correctly ranked lower as they do not directly address enabling notifications or following users. Great job on the perfect ranking!

ContextualRecall: The score is 0.50 because while the nodes in retrieval context support enabling notifications and following pages or users (sentences 1, 3, and 4), they do not mention specific details like 'xwiki.properties', the 'alerte' menu, receiving all user-triggered notifications, or the 'réseau' tab (sentences 2, 5, 6, and 7).

CustomContextualRelevancy: Retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' in 'xwiki.properties' and specifics on following pages or users.

Correctness: The actual output lacks instructions on enabling notifications via 'xwiki.properties', following pages or users, and checking the 'réseau' tab in user profiles. It asks questions instead of providing the expected procedural information.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_023_result.json

Overall_score: 0.6714851280965565

Individual_scores:

AnswerRelevancy: 0.3333333333333333

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Correctness: 0.2955774352460057

Reasons:

AnswerRelevancy: The score is 0.33 because the output includes irrelevant links that do not address the question about enabling notifications for one's own actions.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, ranked first, directly addresses the input question by stating, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter.' The irrelevant nodes, ranked second to fifth, discuss topics such as RSS feeds, general notification settings, links to various settings, and notification preferences macros, none of which specifically address enabling notifications for one's own actions.

ContextualRecall: The score is 0.50 because the first sentence in the expected output is supported by the 1st node in the retrieval context, but the second sentence about disabling the filter in 'Options de

filtrage avancées' is not supported by any node in the retrieval context.

CustomContextualRelevancy: The retrieval context confirms users don't receive notifications for their own actions by default and mentions the option to change this setting, aligning closely with the expected answer. However, it lacks the specific mention of 'Options de filtrage avancées'.

Correctness: The actual output mentions enabling notifications for own actions but omits the default behavior and filtering options section.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_024_result.json

Overall_score: 0.7536657750931447

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Correctness: 0.15381283237705007

Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the question about disabling notifications, but includes irrelevant links that do not directly help in resolving the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than the irrelevant node. The first three nodes in the retrieval context are directly related to modifying notification settings, discussing 'Choosing the default auto-watched mode', 'Create global notification filters', 'Notifications Auto Watch Preferences', and 'Notifications Custom Filters Preferences'. The irrelevant node, ranked fourth, discusses 'RSS feeds and notification macro settings', which are not directly related to the input query about disabling automatic notifications.

ContextualRecall: The score is 0.75 because most sentences in the expected output can be linked to the 2nd node in the retrieval context, which covers topics like 'auto watched pages', 'notification applications', and 'custom filters'. However, one sentence lacks direct support from the retrieval context, leading to a less than perfect score.

CustomContextualRelevancy: The retrieval context includes information about auto-watched pages, custom filters, and notification settings, which align with the expected answer. However, it lacks explicit mention of being notified of changes on your work and modifying the auto-watch setting in user profile notification settings.

Correctness: The actual output does not mention automatic tracking of pages after changes, the ability to modify this setting, or the removal of notification filters as described in the expected output. It focuses on disabling notifications, which contradicts the information in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.35391979160828946

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Correctness: 0.06637589250687943

Reasons:

AnswerRelevancy: The score is 0.86 because the response is mostly relevant to configuring OpenID Connect in XWiki, but it includes a mention of LDAP Authenticator, which is not directly related to the question asked.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'domain-based authentication and session cookies,' which does not pertain to 'automatically activating user accounts using OpenID Connect.' The second node provides details on 'container authentication and LDAP,' which also fails to address 'automatic user activation with OpenID Connect.' The third node, although mentioning OpenID Connect, focuses on 'authentication and synchronization,' not on 'automatically activating user accounts upon first login.' The fourth node lists 'dependencies for the OpenID Connect extension,' unrelated to 'automatic user activation.' Finally, the fifth node is about 'LDAP and Active Directory integration,' which is not connected to 'configuring automatic user activation with OpenID Connect in XWiki.'

ContextualRecall: The score is 0.20 because only the mention of the property `oidc.enableUser` in sentence 1 can be loosely linked to node 1 in the retrieval context, while the rest of the expected output lacks direct support from the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information about modifying the xwiki.properties file or the oidc.enableUser property, which are crucial for the expected answer.

Correctness: The actual output describes configuring `xwiki.authentication.createuser` in `xwiki.cfg`, while the expected output requires updating `oidc.enableUser` in `xwiki.properties`. The steps and files mentioned do not align, leading to a factual discrepancy.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.4300666278454279

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.5137331004059005

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the input question about customizing user profile fields in XWiki, but includes irrelevant statements that provide sources for further reading and introduce a list without directly answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect endpoints' but does not mention 'customizing user profile fields or the `oidc.user.mapping` property.' Similarly, the second node provides 'information on various authentication methods, including OpenID Connect,' but does not address 'customizing user profile fields or mapping properties.' This pattern continues with the third

node, which describes 'the OpenID Connect Authenticator extension' without providing details on 'mapping user profile fields or using `oidc.user.mapping`'. The fourth node lists 'dependencies for the OpenID Connect Authenticator extension,' which is unrelated to 'customizing user profile fields or mapping properties.' Lastly, the fifth node offers 'links to configuration guides and examples' but lacks mention of 'mapping user profile fields or the `oidc.user.mapping` property.'

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention 'xwiki.properties' and OpenID Connect, they lack specific examples of mapping properties, instructions on replacing property names or variables, and details on saving the file and restarting XWiki, which are crucial for full alignment with the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information about customizing user profile fields or using the `oidc.user.mapping` property in `xwiki.properties` as described in the expected output.

Correctness: The actual output covers customization of OpenID Connect fields but lacks specific examples of property mappings and instructions to save and restart XWiki. It introduces correct additional details about default values and the importance of testing.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5274684565311148

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.4444444444444444

CustomContextualRelevancy: 0.3

Correctness: 0.2870329614089118

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 0.80 because the actual output incorrectly claims that group synchronization is enabled by default, while the retrieval context specifies that it requires adding the claim `xwiki_groups` to the property `oidc.userinfo.claims` in the `xwiki.properties` file.

ContextualPrecision: The score is 0.33 because the third node in the retrieval context is relevant, mentioning 'group synchronization, including the use of claims and mapping properties,' which aligns with the input query about configuring group synchronization between XWiki and OpenID Connect. However, the first, second, fourth, and fifth nodes are irrelevant as they discuss 'authentication management and endpoints,' 'different authentication methods,' 'general information about OpenID Connect authentication,' and 'dependencies for the OpenID Connect extension,' respectively, without addressing group synchronization. These irrelevant nodes being ranked higher than the relevant node lowers the score.

ContextualRecall: The score is 0.44 because while the 3rd node in the retrieval context supports configuring group synchronization, including properties like `'oidc.groups.claim'`, `'oidc.groups.separator'`, and `'oidc.groups.mapping'`, it lacks details on locating the `'xwiki.properties'` file, replacing placeholders with actual values, and saving the file and restarting XWiki.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using `oidc.groups.mapping`, but lacks details on configuring `oidc.groups.claim` and `oidc.groups.separator` properties, and saving changes in `xwiki.properties`.

Correctness: Actual output mentions `oidc.userinfoclaims` and automatic group creation, which are not in the expected output. It lacks details on `oidc.groups.separator` and saving the file after changes.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.23955157717651945

Individual_scores:

AnswerRelevancy: 0.375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.062309463059116874

Reasons:

AnswerRelevancy: The score is 0.38 because the output includes multiple irrelevant source links that do not address the question about file upload restrictions, significantly lowering the relevancy. However, the score is not zero, indicating that there is some relevant content present.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'language preferences and translations in XWiki', which is unrelated to file upload restrictions. The second node is about 'various applications and features in XWiki, such as tour guides and administration', but does not mention file uploads or MIME types. The third node covers 'PDF and document processing using XSL and FOP', which does not relate to setting upload restrictions for images or PDFs. The fourth node addresses 'system filters and document viewing', which does not address file upload settings or restrictions. The fifth node contains 'links to various image references' and does not provide information on configuring file upload settings. Lastly, the sixth node is about 'image references and links', which are not relevant to configuring file upload restrictions in a wiki.

ContextualRecall: The score is 0.00 because none of the expected output sentences can be attributed to any node(s) in the retrieval context, as the context lacks information on file size restrictions, MIME type configurations, and specific steps for setting these parameters.

CustomContextualRelevancy: The retrieval context does not contain any information related to configuring MIME types or setting maximum attachment size as outlined in the expected output.

Correctness: The actual output lacks specific procedural steps and does not match factual details like MIME type configuration and exact file size settings present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.5850337980894389

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.010202788536633216

Reasons:

AnswerRelevancy: The score is 0.80 because the main content addresses how to refuse script rights to a space administrator effectively, but the inclusion of repeated source links that do not directly answer the question slightly detracts from the overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because the relevant node in the retrieval context, which discusses 'programming rights can only be granted from the wiki preferences page,' is ranked second. However, the first node is irrelevant as it is about 'language preferences and translations in XWiki,' which does not pertain to script rights or administrative permissions. Additionally, other irrelevant nodes, such as the third node discussing various rights like VIEW, EDIT, COMMENT without addressing script rights, are ranked higher than they should be.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the retrieval context, demonstrating a clear understanding of the topic. Great job!

CustomContextualRelevancy: The retrieval context mentions that admin rights imply other rights and are not deniable, but it does not explicitly state that admin rights cannot be refused, as mentioned in the expected output.

Correctness: The actual output contradicts the expected output by suggesting a method to refuse rights, whereas the expected output states it is not supported.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_030_result.json

Overall_score: 0.3413980325857879

Individual_scores:

AnswerRelevancy: 0.72727272727273

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.3211154682419999

Reasons:

AnswerRelevancy: The score is 0.73 because the output provides some relevant information about configuring spaces and permissions in XWiki, but it includes multiple source links and a note about user groups that do not directly address the specific configuration question, reducing the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node is about 'a specific page in XWiki and does not provide information on setting access rights or managing spaces for teams.' The second node discusses 'modifying resources using XML and HTTP methods, which is not relevant to configuring team access in XWiki.' The third node is about 'managing comments and attachments, which is unrelated to setting access rights for spaces in XWiki.' The fourth node explains 'how spaces and sub-pages work in XWiki, but does not provide guidance on configuring access rights for teams.' The fifth node is about 'migration actions and does not address configuring access rights for teams in XWiki.' Finally, the sixth node discusses 'configuration display options in XWiki, which is not related to setting access rights for team spaces.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to creating groups, setting rights, or managing access in XWiki, which are key elements in the expected output.
Correctness: Actual output creates spaces for teams but misses creating groups for each team and setting wiki-level permissions. It grants read and write permissions instead of specifying modify rights only for team spaces. It lacks mention of inheriting rights for child pages.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_031_result.json

Overall_score: 0.6296716824534421

Individual_scores:

AnswerRelevancy: 0.3333333333333333

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.8

CustomContextualRelevancy: 0.8

Correctness: 0.3446967613873188

Reasons:

AnswerRelevancy: The score is 0.33 because the output contains multiple irrelevant statements that are source URLs, which do not address the question about authenticating users with access tokens. This significantly lowers the relevancy, but the score is not zero, indicating there might be some relevant content present.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because the relevant nodes in the retrieval context are ranked lower than some irrelevant nodes. The first node is about 'LDAP configuration and does not mention token-based authentication or JWTs', which should be ranked lower than the second node that discusses 'Authorized Applications Configuration', relevant to token-based authentication. Similarly, the third node about 'OpenID Connect and OAuth scopes' should be ranked lower than the fourth node, which provides a description of 'authentication using signed JSON Web Tokens (JWT)', directly relevant to the input.

ContextualRecall: The score is 0.80 because most of the expected output sentences are supported by nodes in the retrieval context, particularly node 4 and node 2. However, there is one sentence that lacks direct support from the retrieval context, which slightly lowers the score.

CustomContextualRelevancy: The retrieval context includes information about enabling the token authenticator, configuring authorized applications, and using JWTs, which aligns with the expected answer. However, specific details about JWT claims and the use of Ed25519 keys are not explicitly mentioned.

Correctness: The actual output mentions the use of JWT for authentication but lacks details such as the required claims, Ed25519 signing, and configuration steps mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.25

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.50 because the response partially addresses the topic of authenticating users with access tokens, but includes unrelated source links that detract from the relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Keep up the great work!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'model validation strategies in XWiki', the second node is about 'attachment validation and configuration in XWiki', and the third node provides information on 'the admin UI for model validation'. None of these contexts pertain to user authentication with access tokens, which is why the score is not higher.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context. The retrieval context lacks any mention of key concepts such as 'Benutzer mit Zugriffstoken', 'LLM-Anwendung (BETA)-Erweiterung', 'JSON Web Tokens (JWT)', or the steps and limitations related to using a 'Token-Authentifizierer'.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are key elements in the expected output.

Correctness: The actual output lacks any relevant information regarding access token authentication as described in the expected output, indicating complete factual inaccuracy.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_033_result.json

Overall_score: 0.2807154347744876

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.017625941980259453

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the question about authenticating users with access tokens, but includes irrelevant source links that do not directly contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'upgrading XWiki' without mentioning user authentication or access tokens. The second node talks about 'features of the Administration Application' but lacks details on authenticating users with access tokens. The third node mentions 'managing user rights and profiles' and an authentication guide, yet it does not specifically address the topic of authenticating users with access tokens.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication or JWT, which are key elements in the expected answer.
Correctness: The actual output states it cannot determine if XWiki supports authentication with access tokens, which contradicts the expected output that provides detailed steps on implementing token-based authentication.
Question_language: fr
Expected_answer_language: fr
Answer_language: en

Model: AI.Models.qa_llama3_1_402b

File: qa_001_result.json

Overall_score: 0.32111482221107307

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.03195209116117512

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant and addresses the issue with the 'Bell' icon in XWiki, but the inclusion of source links that do not directly contribute to solving the problem slightly detracts from the overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'installation instructions and prerequisites', the second node describes 'how notifications are displayed', the third node lists 'dependencies for the notifications extension', and the fourth node talks about 'overriding notification templates'. None of these address enabling notifications in the wiki settings.

ContextualRecall: The score is 0.00 because the retrieval context does not contain any information related to enabling notifications in the wiki or the specific settings in `xwiki.properties`, making it impossible to support the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the `notifications.enabled` setting in `xwiki.properties`.

Correctness: The actual output does not address enabling notifications via `xwiki.properties` as the expected output specifies. It focuses on installing the Notifications Application, which is not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.7181391987443496

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.6

Correctness: 0.45883519246609705

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the query about not receiving notifications in XWiki without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.75 because the first node in the retrieval context is relevant, providing information on how to follow a user to receive notifications, which is part of the expected output. However, the second node discusses filters and settings related to notifications but does not directly address enabling notifications or subscribing to pages or users, and should be ranked lower. Similarly, the third node talks about getting a notification RSS feed and settings, which does not provide information on enabling notifications or subscribing to specific content, and should also be ranked lower. The fourth node is relevant as it explains different ways to be notified in XWiki, aligning with the expected output. Lastly, the fifth node is about the 'What's New' application for news updates, which is not directly related to enabling notifications or subscribing to content in XWiki, and should be ranked lower.

ContextualRecall: The score is 0.50 because while some sentences, such as those about using the 'alert' menu (supported by node 4 in retrieval context) and following users (supported by node 1 in retrieval context), are well-supported, key details like enabling notifications in `xwiki.properties` and subscribing to pages or users are not covered by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context covers following users and notification settings but lacks specific details on enabling notifications in xwiki.properties and watching pages.

Correctness: The actual output partially aligns with the expected output by mentioning the need to watch pages or users for notifications. However, it lacks specific instructions about enabling notifications in 'xwiki.properties' and does not cover all methods of watching locations and users described in the expected output. It introduces additional troubleshooting steps not found in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9761317181531796

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.8567903089190769

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question without any irrelevant information. Great job on maintaining focus and clarity!

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job on maintaining perfect alignment!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than irrelevant nodes. The first and second nodes directly address how to enable notifications for one's own actions in XWiki, while the third, fourth, and fifth nodes, which discuss unrelated topics such as administrator configurations, RSS feeds, and general notification methods, are correctly ranked lower. Great job on achieving perfect precision!

ContextualRecall: The score is 1.00 because every part of the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context contains all the information from the expected output, including the default behavior of not receiving notifications about one's own activity unless explicitly targeted, and the ability to change this setting in the 'Advanced filtering options' of the user profile.

Correctness: The actual output accurately describes the default behavior of not receiving notifications for one's own activity and the option to disable this filter under advanced settings, matching the expected output. However, it includes additional steps and version specificity not present in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5463411898849923

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Correctness: 0.228047139309954

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about disabling unwanted notifications in XWiki without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job on ensuring complete alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'getting a notification RSS feed and enabling/disabling notification types', but does not address how to disable notifications or change settings related to automatic page watching. The second node focuses on 'following a user and default watch settings', which also fails to provide information on disabling notifications or changing automatic page watching settings.

Subsequent nodes similarly do not address the query, discussing topics like 'filtering notifications', 'enabling/disabling email notifications', 'administrators configuring notification settings', and 'overriding default notification templates', none of which are relevant to the user's question about disabling notifications in XWiki.

ContextualRecall: The score is 0.75 because while most of the expected output is supported by nodes in the retrieval context, such as the default settings for notifications and the ability to change settings, there is no explicit mention of deleting existing notification filters in the 'Custom Filters' list.

CustomContextualRelevancy: The retrieval context mentions default watch settings and user profile settings but lacks details on automatic page watching and deleting notification filters.

Correctness: The actual output focuses on disabling notifications and unwatching pages, lacking key details like automatic addition of modified pages to watch list, changing 'Automatic page watching' settings, and deleting filters from 'Custom Filters' list.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.824847605665772

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.9

Correctness: 0.815752300661298

Reasons:

AnswerRelevancy: The score is 0.90 because the response is mostly relevant and addresses the question about configuring XWiki for OpenID Connect. However, it includes an irrelevant statement about LDAP authentication, which slightly detracts from its focus.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions the property '# oidc.enableUser=true', is ranked first. This directly addresses the need to enable user accounts on first login using OpenID Connect, as required by the input. The irrelevant nodes, discussing topics like domain-based authentication, general OpenID Connect information, container authentication, and LDAP, are correctly ranked lower.

ContextualRecall: The score is 0.33 because only the first sentence regarding the property '# oidc.enableUser=true' is supported by the 1st node in the retrieval context, while the rest of the instructions and explanations lack direct support from any node.

CustomContextualRelevancy: The retrieval context includes the key information about the 'oidc.enableUser' property, which matches the expected answer's requirement to uncomment and set it to 'true'. However, it lacks explicit instructions on saving the file and restarting XWiki.

Correctness: The actual output correctly identifies the 'oidc.enableUser' property in 'xwiki.properties', specifies setting it to 'true', and explains its purpose. However, it omits the instruction to uncomment the property and the need to save the file and restart XWiki. It also contains additional information about other authentication methods not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.8888232496386848

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6666666666666666

CustomContextualRelevancy: 1.0

Correctness: 0.7496061644987756

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant and mostly addresses the question about customizing user profile fields in XWiki using OpenID Connect. However, it includes a source link that does not directly contribute to answering the specific query, slightly lowering the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node in the retrieval context is directly related to mapping additional user profile fields from the OpenID Connect provider to XWiki user properties, while the subsequent nodes discuss topics

such as access requests for various claims, OpenID Connect provider endpoints, group synchronization, and in-wiki configuration, which are not directly related to the input query.

ContextualRecall: The score is 0.67 because while the main concept of mapping user profile fields from the OpenID Connect provider is supported by the 1st node in the retrieval context, specific instructions about replacing variables and restarting XWiki are not directly supported by any node in the retrieval context.

CustomContextualRelevancy: The retrieval context contains all the necessary information found in the expected answer, including the ability to customize user profile fields using the `oidc.user.mapping` property in the `xwiki.properties` file, with examples provided.

Correctness: The actual output accurately describes the customization of user profile fields using the `oidc.user.mapping` property, consistent with the expected output. However, it lacks details about restarting XWiki and does not use the same example properties and syntax provided in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8462152928167339

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 1.0

Correctness: 0.5772917569004032

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant ones. The first three nodes provide detailed information on configuring properties like `'oidc.groups.claim'`, `'oidc.groups.separator'`, and `'oidc.groups.mapping'`, which are essential for group synchronization. The irrelevant nodes, ranked fourth and fifth, only offer general descriptions about OpenID Connect and authentication management, which do not directly aid in the configuration process.

ContextualRecall: The score is 0.50 because while some sentences in the expected output are supported by nodes in the retrieval context, such as the configuration examples and explanations (nodes 1 and 2), many sentences are general instructions or placeholders that do not directly match any specific part of the retrieval context.

CustomContextualRelevancy: The retrieval context contains all necessary details for configuring group synchronization, including setting the `'oidc.groups.claim'`, `'oidc.groups.separator'`, and `'oidc.groups.mapping'` properties, which align with the expected output instructions.

Correctness: The actual output includes configuration steps for `'oidc.groups.claim'` and `'oidc.groups.mapping'` similar to the expected output but misses details on `'oidc.groups.separator'`. It adds extra configurations like `'oidc.groups.allowed'`, `'oidc.groups.forbidden'`, and `'oidc.groups.prefix'` not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.7046039616491253

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.6363636363636364

CustomContextualRelevancy: 0.5

Correctness: 0.7579268001977821

Reasons:

AnswerRelevancy: The score is 1.00 because the output perfectly addresses the input without any irrelevant statements. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, ranked third, provides information on setting maximum attachment size and mentions mimetype restrictions, which are directly related to the input. However, irrelevant nodes ranked first, second, fourth, and fifth discuss topics such as linking to external files and images, general overview of XWiki's structure, attachment display and security settings, and versioning and storage of attachments, none of which address the configuration of upload size or mimetype restrictions. These irrelevant nodes should be ranked lower than the relevant node.

ContextualRecall: The score is 0.64 because while nodes in the retrieval context support detailed steps for configuring mimetypes and maximum attachment size (as seen in sentences 2-7), introductory and explanatory sentences (1 and 8) and step titles (9 and 10) are not directly supported by the nodes.

CustomContextualRelevancy: The retrieval context mentions configuring maximum attachment size and mimetype restrictions but lacks specific details on setting 10MB limit and mimetype allowList.

Correctness: The actual output correctly mentions setting the maximum upload size to 10MB and configuring mimetype restrictions. However, it introduces the Attachment Validation Application not mentioned in the expected output and uses 'attachment.upload.allowedMimeTypes' instead of 'attachment.upload.allowList'.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.6651530325738324

Individual_scores:

AnswerRelevancy: 0.9375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.05341819544299424

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant to the question about denying script rights to a space administrator, with only a minor irrelevant mention of Nested Pages Migration.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses programming rights but does not address denying script rights to a space administrator. The second node outlines rights but does not mention the inability to deny script rights. The third node details the 'Script' right but does not mention denial to a space administrator. The fourth node talks about access checks, unrelated to denying script rights. The fifth node is about converting rights during migration, not related to denying script rights.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context confirms that rights implied by admin cannot be denied, matching the expected answer's statement about XWiki's lack of support for denying such rights.

Correctness: The actual output incorrectly states that script rights can be denied to a space administrator, while the expected output clearly states that rights implied by admin right cannot be denied in XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7638220862121293

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.6666666666666666

CustomContextualRelevancy: 0.5

Correctness: 0.903445337785597

Reasons:

AnswerRelevancy: The score is 0.85 because the response largely addresses the question about configuring team access in XWiki, but the inclusion of multiple irrelevant source links slightly detracts from its overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Keep up the great work!

ContextualPrecision: The score is 0.67 because the first node in the retrieval context is relevant as it discusses 'rights inheritance and administration features' related to setting view and edit permissions in XWiki. However, the second node, which is ranked higher than it should be, focuses on 'the parent/child relationship and the concept of Nested Pages,' which does not directly address configuring view/edit rights for teams. Additionally, the third node, about 'string formatting and user identification,' is unrelated to configuring access rights and should be ranked lower than the relevant nodes.

ContextualRecall: The score is 0.67 because while nodes in the retrieval context provide support for detailed steps such as creating groups, setting wiki-wide and space-level rights, and inheriting rights for child pages, introductory and summary statements in the expected output do not directly align with specific nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions rights inheritance and setting rights at different levels, which aligns partially with the expected answer's steps on setting wiki-wide and space-level rights. However, it lacks specific details about creating groups for each team and setting the 'Edit' right specifically for team spaces.

Correctness: The actual output closely matches the expected output with slight variations in wording. Both outline creating groups, setting view rights at the wiki level, setting edit rights at the space level,

and ensuring rights inheritance, although 'XWikiAllGroup' is not explicitly mentioned in the actual output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.46151747957511585

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.22743821078402865

Reasons:

AnswerRelevancy: The score is 0.88 because the response is mostly relevant to the issue of the unavailable bell icon, but it includes source and documentation links that do not directly address the specific problem mentioned.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node in retrieval context, ranked sixth, mentions 'editing the xwiki.properties file', which is pertinent to enabling notifications. However, the first five nodes in retrieval context are irrelevant as they discuss 'custom notification filters', 'alert menu', 'different ways to receive notifications', 'links to various notification-related topics', and 'notification email settings', none of which address enabling notifications through the xwiki.properties file.

ContextualRecall: The score is 0.50 because while the 5th node in the retrieval context mentions editing the xwiki.properties file, there is no specific mention of setting 'notifications.enabled' to 'true', which is crucial for fully supporting the expected output's instructions.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via 'xwiki.properties' or setting 'notifications.enabled' to 'true'.

Correctness: The actual output does not mention enabling notifications via 'notifications.enabled' in 'xwiki.properties'. It discusses unrelated factors for notification visibility, which are factually correct but irrelevant to the expected task.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.330141126151344

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.14751342357473066

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the issue of not receiving notifications, but it includes irrelevant information about XWiki-Skins, which slightly detracts from its

overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node contains 'a series of image links and references which do not relate to the topic of enabling notifications or following users/pages', and similar issues persist in the subsequent nodes. This results in irrelevant nodes being ranked higher than any relevant content.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context contains image links and references, which do not provide any information related to enabling notifications or following pages and users as described in the expected output.

Correctness: The actual output mentions checking notification settings but lacks details on following pages or users and setting `notifications.enabled` to `true`. It also includes unrelated information about spam folders and contacting administrators, which is not in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.33294248926254355

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.1794731173934429

Reasons:

AnswerRelevancy: The score is 0.82 because the response is mostly relevant to the question about enabling notifications for one's own actions, but it includes source links that do not directly address the specific query, slightly detracting from the overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customizing page templates, internationalization, accessibility, and scalability', which are unrelated to notification settings. The second node provides information about 'administrating wiki features and the XWiki Development Team', which does not relate to enabling notifications for one's own activity. The third node mentions 'various XWiki extensions like Office Importer and Syntax Highlighting', but none are relevant to notification settings or user activity.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters mentioned in the expected output.

Correctness: The actual output does not align with the key facts of the expected output, lacks critical details about disabling the filter in 'Erweiterten Filteroptionen', and introduces extraneous steps not mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.3041238795159346

Individual_scores:

AnswerRelevancy: 0.6875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.13724327709560757

Reasons:

AnswerRelevancy: The score is 0.69 because while the response contains some relevant information about disabling notifications, it is diluted by multiple irrelevant links to XWiki pages that do not directly address the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files', which does not relate to 'notification settings or disabling notifications'. The second node talks about 'moving and deleting attachments', which is unrelated to 'managing or disabling notifications'. The third node is about 'displaying content and macros', which does not address 'notification settings'. The fourth node involves 'administrating wiki features', which does not provide information on 'notification settings'. The fifth node discusses 'permissions and security modules', which are not related to 'notification settings'. Finally, the sixth node lists 'various applications and features', none of which pertain to 'notification settings or disabling notifications'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about automatic page following, notification settings, or custom filters as mentioned in the expected output.

Correctness: The actual output does not address the specific feature of automatically tracking pages after significant changes, nor does it mention modifying notification settings or deleting filters, which are key points in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.4324332816261954

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.1945996897571724

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring XWiki for automatic user account activation with OpenID Connect. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'domain-based authentication and session cookie configuration',

which is unrelated to 'automatic activation of user accounts upon first login with OpenID Connect'. The second node provides 'general information about OpenID Connect and its integration with XWiki', but lacks details on the specific configuration needed for automatic user activation. The third node describes 'OpenID Connect authenticator and its compatibility' without addressing the automatic user activation configuration. Lastly, the fourth node mentions 'xwiki.properties file' but focuses on 'endpoints and provider configuration', not on enabling automatic user activation with OpenID Connect.

ContextualRecall: The score is 0.40 because while the 4th node in the retrieval context supports the mention of 'xwiki.properties' and OpenID Connect, it lacks details on updating specific properties, saving the file, restarting XWiki, and automatic activation of new user accounts.

CustomContextualRelevancy: The retrieval context does not mention adjusting the xwiki.properties file or the specific property `oidc.enableUser`, which are crucial for the expected output.

Correctness: The actual output acknowledges the lack of direct documentation on automatic activation, missing the ``xwiki.properties`` file and ``oidc.enableUser=true`` property details from the expected output. It provides general advice but omits key configuration steps, lacking specific instructions and contradicting the expected clarity and precision.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.4607123663801787

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.3642741982810726

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect endpoints' without mentioning 'user profile field mapping or customization'. The second node provides information on 'container authentication and other authentication methods' but does not address 'mapping user profile fields from an OpenID Connect provider'. The third node offers a 'general description of the OpenID Connect Authenticator and its configuration guides' without mentioning 'customizing user profile fields'. The fourth node describes the 'OpenID Connect Authenticator extension's details and compatibility' but does not address 'user profile field mapping'. Finally, the fifth node lists 'dependencies for the OpenID Connect Authenticator extension' but does not mention 'user profile field customization or mapping'.

ContextualRecall: The score is 0.40 because while the nodes in retrieval context mention 'xwiki.properties' and ``oidc.user.mapping``, which support the customization of user profile fields (sentences 1 and 2), they lack specific examples and instructions found in sentences 3 and 4, such as mapping custom properties, replacing variables, and restarting XWiki.

CustomContextualRelevancy: The retrieval context lacks any information about customizing user profile fields using ``oidc.user.mapping`` in ``xwiki.properties`` as described in the expected output.

Correctness: The Actual Output acknowledges customization of user profile fields from OpenID Connect, aligning with the Expected Output. However, it lacks specific configuration details such as ``xwiki.properties`` and ``oidc.user.mapping`` mentioned in the Expected Output, suggesting editing

`xwiki.cfg` instead, which is inconsistent.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.40300017610084454

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.16666666666666666

CustomContextualRelevancy: 0.0

Correctness: 0.25133438993840057

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about setting up group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'general OpenID Connect provider configuration and endpoints' without addressing group synchronization. The second node talks about 'container authentication and other authentication methods like LDAP,' which is unrelated to the setup of group synchronization with OpenID Connect. The third node provides 'links to configuration guides' but lacks specific information on group synchronization. The fourth node lists 'dependencies for the OpenID Connect authenticator' without mentioning group synchronization. Finally, the fifth node covers 'general information about the OpenID Connect authenticator' but not the setup for group synchronization.

ContextualRecall: The score is 0.17 because only the configurations for `oidc.groups.claim` and `oidc.groups.separator` mentioned in sentence 3 and 4 of the expected output are supported by the 1st node in the retrieval context. Most other instructions and configurations, such as finding and editing the `xwiki.properties` file, setting up group synchronization, and mapping groups, are not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks specific instructions on configuring 'xwiki.properties' for group synchronization, such as setting 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping'.

Correctness: The actual output lacks specific configuration steps mentioned in the expected output, such as editing 'xwiki.properties' and setting properties like 'oidc.groups.claim'. It provides general guidance instead of detailed instructions.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6639883998345474

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.325

ContextualRecall: 0.5454545454545454

CustomContextualRelevancy: 0.5
Correctness: 0.6134758535527389

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the input without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the first three nodes in the retrieval context are irrelevant, discussing 'trusted domains and PDF export customization', 'ODT and PDF export features', and 'various features and macros', none of which relate to configuring attachment upload restrictions. The relevant nodes, which mention 'Size Limit' and 'Mimetype Restriction' and provide 'detailed steps on setting the maximum size of an attachment', are ranked lower than these irrelevant nodes.

ContextualRecall: The score is 0.55 because while there is detailed support in the nodes in retrieval context for steps related to configuring maximum attachment size in XWikiPreferences (as seen in node 5), there is a lack of information regarding the configuration of MIME types and the specific line for MIME type restrictions, which are not covered in the nodes.

CustomContextualRelevancy: The retrieval context partially covers the expected answer by mentioning attachment size limits and MIME type restrictions, but lacks specific details on configuring MIME types in `xwiki.properties` and setting the exact size to 10 MB.

Correctness: The actual output correctly limits uploads to image and PDF files and sets the max size to 10 MB. However, it suggests using the Attachment-Validation-Application instead of editing ``xwiki.properties``, and does not mention repeating the size configuration for each (sub)wiki.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.33521391078658086

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.16512961856563893

Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses the question about denying script rights to a domain administrator, but includes irrelevant source links that do not directly contribute to answering the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'form validation' without mentioning script rights or administrative permissions, the second node is about 'templates and page creation' which is unrelated, and the third node talks about 'configuration classes and custom rights checks' but does not address denying script rights to a space administrator.

ContextualRecall: The score is 0.00 because the sentence about XWiki's admin rights is not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about XWiki not supporting the denial of rights implied by admin rights.

Correctness: The actual output does not align with the expected output as it suggests steps for denying script rights, which contradicts the expected message that such denial is unsupported in XWiki. There are also additional incorrect instructions present.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.4458142830484104

Individual_scores:

AnswerRelevancy: 0.782608695652174

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2222222222222222

CustomContextualRelevancy: 0.0

Correctness: 0.6700547804160664

Reasons:

AnswerRelevancy: The score is 0.78 because the response partially addresses the configuration of access rights in XWiki, but includes multiple irrelevant source links that do not directly relate to the specific question about configuring team access and editing permissions.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses XWiki's database support and programming features, which are unrelated to configuring access control for teams. The second node is about the Notifications Application, not related to access controls. The third node explains Nested Pages and subwikis, which do not address team-specific access controls. The fourth node provides information on managing users and groups but does not specifically address configuring access controls for team areas. Finally, the fifth node discusses group synchronization and customization, which is not directly relevant to the specific access control configuration needed.

ContextualRecall: The score is 0.22 because only sentence 1 and sentence 4 of the expected output have supportive reasons linked to nodes in the retrieval context, specifically nodes 4 and 3 respectively. The rest of the sentences lack corresponding information in the nodes, leading to a low score.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting wiki-wide or area-specific rights, or inheriting rights for subpages as outlined in the expected output.

Correctness: Actual output aligns with the expected outcome by ensuring all teams can view all areas and only edit their own. However, it lacks details on setting wiki-wide rights, inheriting rights for child pages, and emphasizes creating a separate area for each team, which is not specified in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.3558472322331614

Individual_scores:

AnswerRelevancy: 0.9411764705882353

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.19390692281073318

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant to the issue of the notification bell not being available, but it includes an unnecessary link to XWiki syntax documentation, which slightly detracts from its focus.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'custom notification filters and their behavior but does not provide information on enabling notifications in the wiki settings.' Similarly, the second node 'lists various notification settings and customization options but does not mention enabling notifications through the `xwiki.properties` file.' All subsequent nodes also fail to address the issue of enabling notifications in the wiki, which is the core of the input query.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be linked to any nodes in the retrieval context, indicating a complete mismatch between the expected information and the available context.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the 'xwiki.properties' file or setting 'notifications.enabled' to 'true'.

Correctness: The actual output does not address enabling notifications via `notifications.enabled` in `xwiki.properties` as in the expected output. It instead focuses on troubleshooting the notification display issue and includes unrelated details about checking extensions and permissions.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.6434575003772078

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Correctness: 0.22741166892991388

Reasons:

AnswerRelevancy: The score is 0.83 because the response is mostly relevant to the issue of not receiving notifications, but it includes a source link and a link to XWiki syntax documentation that do not directly address the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant nodes in the retrieval context are perfectly ranked higher than the irrelevant nodes. The first node discusses 'Enable and disable the extension' and 'Enabling notifications for all users by default', which directly relates to ensuring notifications are activated. The second node includes 'Follow a user', which is relevant to the guidance on following users to receive notifications. The irrelevant nodes, ranked third and fourth, discuss topics like 'getting a notification RSS feed and settings' and 'various notification preferences macros', which do not directly address the specific actions needed to ensure notifications are received.

ContextualRecall: The score is 0.50 because while nodes in the retrieval context support the general concepts of enabling notifications and following users or pages (as seen in sentences 1, 3, and 5 of the

expected output), they lack specific details such as setting 'notifications.enabled' in 'xwiki.properties', the 'alerte' menu, and the 'réseau' tab mentioned in sentences 2, 4, and 6.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' to true and specifics on following pages or users as described in the expected output.

Correctness: The actual output lacks specific instructions on setting 'notifications.enabled' in 'xwiki.properties' and following pages or users for notifications, which are critical in the expected output. It focuses more on general settings and troubleshooting, which is inconsistent with the detailed steps outlined in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8201409823869619

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Correctness: 0.854179227655105

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the question about enabling notifications for one's own actions, but includes irrelevant source links that do not contribute to solving the issue.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which states, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' is ranked first. This directly addresses the input query about default notification settings and how to change them. The irrelevant nodes, which discuss topics like notification RSS feeds, lists of links, and various technical aspects of notifications without addressing the specific issue of receiving notifications for one's own actions, are all ranked lower.

ContextualRecall: The score is 0.50 because the first sentence aligns with the information in the 1st node in the retrieval context, but the second sentence about disabling the filter in 'Options de filtrage avancées' is not supported by any node in the retrieval context.

CustomContextualRelevancy: The retrieval context confirms that by default, users do not receive notifications for their own actions and mentions the ability to change this setting, aligning closely with the expected output. However, it does not explicitly mention the 'Options de filtrage avancées' section.

Correctness: Actual Output accurately describes the default notification settings and mentions the ability to disable the filter, aligning with Expected Output, but uses different wording and lacks specific mention of 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.8075314683231044

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Correctness: 0.5174110321608492

Reasons:

AnswerRelevancy: The score is 0.78 because the response mostly addresses the question about disabling notifications, but it includes irrelevant source links that do not directly help with the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant node. The first three nodes effectively address modifying 'Suivi automatique des pages' and managing notification settings, while the fourth node, discussing 'RSS feeds and notification macros', is correctly ranked lower as it does not pertain to disabling automatic notifications.

ContextualRecall: The score is 0.75 because while nodes in the retrieval context support the mention of 'auto watched pages', 'Suivi automatique des pages', and 'Filtres personnalisés', there is no reference to being notified of changes on your work.

CustomContextualRelevancy: The retrieval context includes information about auto-watched pages, custom notification filters, and user notification preferences, which aligns with the expected output about automatic page following and notification settings. However, it lacks explicit mention of modifying the 'Suivi automatique des pages' setting and removing filters in 'Filtres personnalisés'.

Correctness: The actual output discusses 'Auto Watch' similar to the expected output's 'Suivi automatique des pages' but lacks mention of 'Filtres personnalisés' for notification settings. It provides steps to manage notifications, aligning partially with the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.3595175208201478

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 0.8

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.1878743556901176

Reasons:

AnswerRelevancy: The score is 0.77 because the response addresses the main question about configuring OpenID Connect for automatic user activation, but it includes irrelevant details about LDAP authentication, which are not related to the query.

Faithfulness: The score is 0.80 because the actual output incorrectly suggests setting `xwiki.authentication.createuser` to `empty` for automatic user account creation, while the retrieval context only specifies that it must be set in xwiki.cfg, without mentioning the `empty` value.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'domain-based authentication and session cookies', which is unrelated to the automatic enabling of user accounts with OpenID Connect. The second node focuses on 'container authentication' and mentions OpenID Connect but lacks information on enabling user accounts automatically. The third node describes the 'OpenID Connect Authenticator extension' without addressing the automatic enabling of user accounts. The fourth node lists 'dependencies for the

OpenID Connect extension', which does not pertain to the required configuration. Lastly, the fifth node is about 'LDAP and Active Directory integration', which is unrelated to the input query.

ContextualRecall: The score is 0.40 because while nodes in the retrieval context mention 'OpenId Connect Authentication' and 'xwiki.properties', they do not specifically address adding or updating the 'oidc.enableUser' property, saving the file, or restarting XWiki, which are crucial steps mentioned in the expected output.

CustomContextualRelevancy: The retrieval context does not mention adjusting the xwiki.properties file or the specific property oidc.enableUser for automatic user account activation via OpenID Connect.

Correctness: The actual output suggests modifying `xwiki.cfg` with `xwiki.authentication.createuser=empty`, whereas the expected output requires modifying `xwiki.properties` with `oidc.enableUser=true`. The actual output also includes irrelevant information about LDAP and version variability.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.40715135419269277

Individual_scores:

AnswerRelevancy: 0.7272727272727273

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.3156353978834291

Reasons:

AnswerRelevancy: The score is 0.73 because the response partially addresses the question by mentioning sources and documentation related to OpenID Connect in XWiki. However, it includes irrelevant details that do not directly explain how to customize user profile fields, which prevents a higher score.

Faithfulness: The score is 1.00 because there are no contradictions; the actual output aligns perfectly with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect provider configurations' but lacks information on 'mapping user profile fields'. The second node is about 'container authentication and other authentication methods like LDAP and Active Directory', which do not pertain to 'mapping OpenID Connect user profile fields'. The third node provides details on the 'OpenID Connect Authenticator extension' but does not cover 'user profile field mapping'. The fourth node lists 'dependencies for the OpenID Connect Authenticator extension', which is unrelated to the topic of 'mapping user profile fields'. Lastly, the fifth node describes 'configuration guides and limitations for OpenID Connect' without mentioning 'user profile field mapping'.

ContextualRecall: The score is 0.40 because while the nodes in retrieval context mention 'xwiki.properties' and 'oidc.user.mapping', aligning with the customization of user profile fields in the expected output, they lack specific examples of mapping properties, instructions on replacing placeholders with actual property names, and the process of saving and restarting XWiki to apply changes.

CustomContextualRelevancy: The retrieval context lacks information on customizing user profile fields using `oidc.user.mapping` in `xwiki.properties`, which is the main focus of the expected output.

Correctness: The actual output mentions the possibility of customizing user profile fields from OpenID Connect but lacks specific details about mapping in `xwiki.properties` and the use of `oidc.user.mapping`. It also does not provide example mappings or mention restarting XWiki, as

specified in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5926009651824126

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.4444444444444444

CustomContextualRelevancy: 0.5

Correctness: 0.27782801331669804

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider, with no irrelevant statements.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which 'directly mentions group synchronization, including the use of claims and mapping properties,' is ranked third. The first and second nodes, which 'do not address group synchronization or the specific properties needed for it' and 'do not provide details on group synchronization configuration,' are ranked higher than the relevant node. This lowers the score since irrelevant nodes are prioritized over the relevant one.

ContextualRecall: The score is 0.44 because while the retrieval context nodes support the general concept of group synchronization and configuration in XWiki, they lack specific details such as opening the file with a text editor, replacing placeholder values, and saving and restarting XWiki, which are crucial for a complete understanding.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using `oidc.groups.mapping`, but lacks details on configuring `oidc.groups.claim` and `oidc.groups.separator`, which are present in the expected output.

Correctness: Actual output introduces properties ``oidc.userinfoclaims`` and ``oidc.groupsync.enabled`` not found in expected output, lacks details on configuring ``oidc.groups.claim`` and ``oidc.groups.separator``.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3567846593888285

Individual_scores:

AnswerRelevancy: 0.6428571428571429

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.497850813475828

Reasons:

AnswerRelevancy: The score is 0.64 because the response partially addresses the question about configuring file upload restrictions, but it includes multiple irrelevant source links that do not directly address the specific requirements of allowing only image or PDF files up to 10 MB.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For instance, the first node discusses 'language preferences and translations in XWiki', which is unrelated to file upload restrictions. Similarly, the second node provides information about 'various XWiki applications and their functionalities', but does not address file upload restrictions. None of the nodes relate to setting file upload restrictions, resulting in a low score.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context. Each sentence either introduces a concept or provides specific instructions that are not covered in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to configuring MIME types or setting attachment size limits as outlined in the expected output.

Correctness: The actual output allows specific image types and mentions server restart, but misses the MIME wildcard and farm-level configuration. The max size configuration is correct but lacks specific guidance on UI steps as in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.616132708102611

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.2467962486156655

Reasons:

AnswerRelevancy: The score is 0.67 because the answer contains some relevant information about script rights, but it includes irrelevant source links that do not directly address the specific question about refusing script rights to a space administrator.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.58 because relevant nodes discussing programming and script rights in XWiki (second node) and detailed information on rights like 'ADMIN' and 'SCRIPT' (third node) are ranked lower than irrelevant nodes. The first node discusses language preferences and translations in XWiki, which is unrelated to script rights or administrative rights, and should be ranked lower. Additionally, the fourth node contains technical details about checking access rights in the code, which does not directly address the question, and the fifth node discusses executing code with certain user rights, which is not relevant to refusing script rights to an administrator.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the information in the nodes in retrieval context, showing a clear understanding of XWiki's rights and permissions. Great job!

CustomContextualRelevancy: The retrieval context mentions that ADMIN rights are not deniable, which partially aligns with the expected answer about administrator rights not being revocable. However, it lacks specific mention of XWiki not supporting this feature as stated in the expected output.

Correctness: The actual output suggests revoking the 'Script' right from a space admin, which contradicts the expected output stating that admin rights cannot be refused in XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.40641629154722786

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.6607199715055896

Reasons:

AnswerRelevancy: The score is 0.78 because the output provides some relevant information regarding configuring access rights in XWiki, but it includes several irrelevant links that do not directly address the question about setting permissions for team spaces.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'the structure and metadata of a page in XWiki', which does not relate to configuring access rights for teams. The second node talks about 'modifying resources using XML and HTTP methods', which is not relevant to setting access controls in XWiki. The third node is about 'adding, editing, and managing comments', which does not pertain to configuring team-specific access in XWiki. The fourth node explains 'the concept of spaces and pages in XWiki', but does not provide information on setting access rights for teams. The fifth node lists 'actions needed after running a migrator', which does not address configuring access rights for teams in XWiki. Lastly, the sixth node describes 'configuration display scopes in XWiki', which does not directly relate to setting access rights for teams.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, defining rights at the wiki or space level, or inheriting rights for child pages as described in the expected output.

Correctness: The actual output captures key steps like creating groups and setting access rights but lacks details on inheriting rights for child pages. It covers assigning read access to 'XWikiAllGroup' but doesn't specify 'Page & Enfants' inheritance.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7128582034204819

Individual_scores:

AnswerRelevancy: 0.8260869565217391

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8
Correctness: 0.43677654971543767

Reasons:

AnswerRelevancy: The score is 0.83 because the main content effectively addresses the question about authenticating users with access tokens, but the presence of multiple irrelevant source link statements prevents the score from being higher.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.50 because the second node in the retrieval context is relevant, describing the configuration of 'Authorized Applications,' which pertains to using tokens for authentication, and the fourth node is also relevant, providing a description of the extension for authentication with JWTs. However, the first node, ranked highest, is irrelevant as it focuses on LDAP configuration without mentioning token-based authentication, and the third node discusses OpenID Connect scopes, which are not directly related to token-based authentication with JWTs. Additionally, the fifth node talks about general authentication mechanisms without specifically addressing token-based authentication.

ContextualRecall: The score is 0.71 because while several aspects of the expected output are supported by the nodes in the retrieval context, such as the use of JWT for authentication and configuration of authorized applications, there are missing details like generating a JWT token with specific claims and including it in the 'Authorization' header.

CustomContextualRelevancy: The retrieval context includes information about enabling the token authenticator, configuring authorized applications, and using JWTs, which aligns with the expected answer. However, details about specific JWT claims and the use of Ed25519 keys are missing.

Correctness: The actual output mentions LLM Application Authenticator but lacks details on JWT claims, signing, and limitations noted in expected output. It also discusses OIDC and custom authenticators, which are irrelevant.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.2955589575042842

Individual_scores:

AnswerRelevancy: 0.7272727272727273

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.04608101775297775

Reasons:

AnswerRelevancy: The score is 0.73 because the response partially addresses the question about authenticating users with access tokens, but includes irrelevant information about source links and unrelated topics like XWiki platform and its extensions.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'Model Validation in XWiki', which is unrelated to 'user authentication with access tokens'. The second node is about 'attachment validation and configuration', not related to 'authentication using access tokens'. The third node provides details on the 'Admin UI for Model Validation', which does not relate to 'authentication methods or access tokens'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, indicating a complete mismatch between the expected content and the available information.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are key aspects of the expected answer.

Correctness: Actual output lacks alignment with expected output, missing specific details about JWT authentication and configuration steps in XWiki.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.32154485497187374

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.11108731164942429

Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the question about authenticating users with access tokens, but it includes source links that are not directly related to the topic, slightly reducing its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'upgrading XWiki' without mentioning 'user authentication or token-based authentication'. The second node talks about 'features related to users and rights' but does not 'specifically address token-based authentication or JWT'. The third node mentions 'rights management' and links to 'authentication guides', yet it does not 'directly address the use of access tokens for authentication'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, indicating a complete lack of relevant information.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are central to the expected output.

Correctness: Actual output lacks mention of JWT, Ed25519, xwiki.authentication.authclass, and specifics about configuring applications, which are key facts in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_llama3_1_8b_Q4

File: qa_001_result.json

Overall_score: 0.3321727607743676

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.14688271849235918

Reasons:

AnswerRelevancy: The score is 0.85 because the response largely addresses the issue of the Bell icon not being visible in XWiki, but it includes irrelevant source links that do not directly help in resolving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question about enabling notifications in xwiki.properties. The first node discusses 'installation instructions and prerequisites', which do not relate to the issue at hand. The second node describes 'the functionality of the alert menu and notifications', but fails to address the enabling of notifications. The third node lists 'dependencies for the extension', which is unrelated to the problem. Finally, the fourth node talks about 'overriding notification templates', which is also not relevant. Therefore, no relevant nodes are ranked higher than irrelevant ones.

ContextualRecall: The score is 0.00 because the retrieval context does not provide any information about enabling notifications in the wiki or the specific settings in `xwiki.properties`.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the `notifications.enabled` setting in `xwiki.properties`.

Correctness: The actual output does not mention enabling notifications in xwiki.properties, which is required in the expected output, focusing instead on troubleshooting missing Bell icon and dependencies.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6255948333888044

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.6666666666666666

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Correctness: 0.33690233366615974

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the issue of not receiving notifications in XWiki. Great job!

Faithfulness: The score is 0.67 because the actual output incorrectly claims that new users don't watch anything or enable notification types by default, whereas the retrieval context clarifies that in XWiki versions before 15.5, new users have the whole wiki watched by default, and in versions 15.5 and later, all application types are enabled by default.

ContextualPrecision: The score is 0.75 because the first and fourth nodes in the retrieval context are relevant, addressing how to receive notifications in XWiki, and are ranked higher. However, the second node, which discusses 'filters and settings related to notifications but does not directly address enabling notifications or subscribing to content', and the third node, which 'describes obtaining an RSS feed for notifications, which is not directly related to enabling notifications or subscribing to content', are ranked higher than the fourth relevant node. Additionally, the fifth node, which 'describes the 'What's New' application, which is not directly related to enabling notifications or subscribing to content', is ranked lower, which is correct.

ContextualRecall: The score is 0.50 because while the retrieval context covers aspects like enabling notifications through the 'alert' menu, following users, and using the 'network' tab (as seen in node 1), it lacks information about enabling notifications in `xwiki.properties`, subscribing to pages or users, and watching a page or its children.

CustomContextualRelevancy: The retrieval context covers user following and notification settings but lacks explicit mention of enabling notifications via `xwiki.properties` or watching pages and wikis.

Correctness: The actual output mentions watching users and enabling notifications but lacks details on enabling notifications in 'xwiki.properties' and the specific process of subscribing to pages or users. It also does not mention the 'alert' menu and 'network' tab.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9455944490710917

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.6735666944265505

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about enabling notifications for one's own actions in XWiki. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than irrelevant nodes. The first node clearly states, 'By default, one don't receive notification about one's own activity... You have the ability to disable this filter in the "Advanced filtering options" section of your own notification settings,' which directly addresses the input. The second node also supports this by explaining how to enable notifications for one's own actions. Irrelevant nodes, such as the third node, which is about administrators configuring notifications, are ranked lower.

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context contains all the information from the expected output, including the default setting of not receiving notifications about one's own activity unless explicitly targeted, and the ability to disable this filter in the 'Advanced filtering options' section of the user's notification settings.

Correctness: The actual output provides a method to enable notifications for one's own activity, aligning with the expected output's mention of this feature. However, it incorrectly states modifying the xwiki.properties file as an alternative, which is not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5339119390513982

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 0.8

ContextualPrecision: 0.2

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Correctness: 0.23680496764172276

Reasons:

AnswerRelevancy: The score is 0.92 because the answer effectively addresses the question about disabling notifications, but includes an unnecessary source link that doesn't directly pertain to the solution.

Faithfulness: The score is 0.80 because the actual output incorrectly suggests that users can disable unwanted notifications themselves, while the retrieval context specifies that only administrators have the ability to change users' notification settings in XWiki 13.2+.

ContextualPrecision: The score is 0.20 because the relevant node in the retrieval context, which is the fifth node, is ranked lower than several irrelevant nodes. The first node discusses 'getting a notification RSS feed and enabling/disabling notification types', which does not address disabling notifications for uninterested pages. The second node talks about 'following a user and default watch settings', which is not about disabling notifications for specific pages. The third node describes 'filtering notifications for hidden pages and minor events', which is also not relevant. The fourth node is about 'enabling/disabling email notifications and customizing email notification settings', which does not directly relate to the input query. The sixth node discusses 'overriding default notification templates and installation instructions', which is not relevant to the input. The relevant node, which should be ranked higher, mentions 'administrators can change notification settings and default auto-watched mode', directly relating to changing settings for automatic page watching.

ContextualRecall: The score is 0.75 because most aspects of the expected output are supported by nodes in the retrieval context, such as automatic page watching and notification settings. However, there is no specific mention of deleting existing notification filters in the 'Custom Filters' list, which affects the score.

CustomContextualRelevancy: The retrieval context mentions default notification settings and the ability to change them, but lacks details about automatic page watching and deleting filters from the 'Custom Filters' list.

Correctness: Actual output suggests disabling notifications via toggling settings, diverging from expected output, which focuses on 'Automatic page watching' and 'Custom Filters'.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.829918931861671

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 0.8333333333333334

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.9

Correctness: 0.7295135911700258

Reasons:

AnswerRelevancy: The score is 0.92 because the response is mostly relevant and helpful in addressing the OpenID Connect configuration issue, but it includes an unnecessary reference to LDAP authentication, which slightly detracts from its focus.

Faithfulness: The score is 0.83 because the actual output incorrectly states that OIDC authentication does not automatically enable user accounts upon their first login by default, whereas the retrieval

context indicates that user accounts should indeed be enabled on first login by default in OIDC configuration.

ContextualPrecision: The score is 1.00 because the relevant node, which mentions the property 'oidc.enableUser=true' directly related to enabling user accounts on first login, is ranked first. Irrelevant nodes, such as those discussing domain-based authentication and session cookies (second node), general OpenID Connect information (third node), container authentication and Active Directory (fourth node), and LDAP configuration (fifth node), are ranked lower. Great job on the perfect ranking!

ContextualRecall: The score is 0.60 because while the retrieval context nodes provide information on defining and setting the 'oidc.enableUser' property (related to sentences 1, 2, and 3), they lack details on saving the file, restarting XWiki, and the automatic enabling of user accounts (related to sentences 4 and 5).

CustomContextualRelevancy: The retrieval context contains the relevant information about the 'oidc.enableUser' property, but lacks explicit instructions to uncomment and set it to 'true', save the file, and restart the XWiki instance.

Correctness: The actual output correctly mentions setting 'oidc.enableUser' to 'true' to enable accounts at first login, aligning with the expected output's key fact. However, it refers to modifying '/WEB-INF/xwiki.cfg' instead of 'xwiki.properties', and lacks mention of restarting the XWiki instance, which are discrepancies.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.921017685071977

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 0.8888888888888888

ContextualPrecision: 1.0

ContextualRecall: 0.8

CustomContextualRelevancy: 1.0

Correctness: 0.908645792971545

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to the input question about customizing user profile fields in XWiki using OpenID Connect. However, it includes a source reference that, while informative, does not directly address the customization query, slightly affecting the score.

Faithfulness: The score is 0.89 because the actual output claims a specific mapping method 'oidc.user.mapping=myxproperty1=\${oidc.user.subject}' which is not mentioned in the retrieval context, indicating a slight misalignment.

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which includes information about mapping non-standard properties from the OpenID Connect provider to XWiki user properties, is ranked first. This directly addresses the input query about customizing user profile fields.

ContextualRecall: The score is 0.80 because the retrieval context supports most of the expected output, particularly the customization of user profile fields and mapping examples, as seen in the 1st node in retrieval context. However, it lacks explicit information about saving the file and restarting XWiki.

CustomContextualRelevancy: The retrieval context contains all information from the expected answer, including customization of user profile fields via the oidc.user.mapping property in xwiki.properties and examples of mapping custom properties.

Correctness: The actual output accurately reflects the expected output, covering customization of user profile fields using `oidc.user.mapping`. It provides examples similar to those in the expected output. The additional information about multiple mappings adds value without factual errors. The mention of `oidc.user.nameFormatter` is extra but factually correct, causing a minor deduction.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8271119936311105

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.46153846153846156

CustomContextualRelevancy: 0.9

Correctness: 0.6011335002482012

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than irrelevant nodes. The first three nodes provide detailed information about configuring the `oidc.groups.claim` and `oidc.groups.mapping` properties, which are essential for synchronizing groups between XWiki and an OpenID Connect provider. The irrelevant nodes, ranked fourth and fifth, only offer general information about OpenID Connect and authentication management, which do not directly address group synchronization.

ContextualRecall: The score is 0.46 because while nodes in the retrieval context provide detailed information on configuring properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` (supporting sentences 3, 4, and 5), they lack instructions for file handling and specific replacements (unsupporting sentences 1, 2, 6, and 7).

CustomContextualRelevancy: The retrieval context contains detailed information about configuring group synchronization, including setting the `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` properties, which aligns closely with the expected output. However, it lacks explicit instructions on locating and editing the `xwiki.properties` file and restarting the XWiki instance.

Correctness: The actual output includes key steps like setting `oidc.groups.claim` and using `oidc.groups.mapping`, but lacks specific instruction on opening and saving the `xwiki.properties` file, and omits the example claim and separator details.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.5688424645805416

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 0.8571428571428571

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.6363636363636364

CustomContextualRelevancy: 0.3

Correctness: 0.45288162731008985

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the question about restricting file types and sizes for uploads, but includes some irrelevant information about XWiki basic concepts and attachment versioning, which are not directly related to the specific requirement.

Faithfulness: The score is 0.86 because the actual output incorrectly suggests that attachment versioning can be disabled in XWiki by setting `xwiki.store.attachment.versioning=0`, whereas the retrieval context indicates that XWiki stores all document attachment versions by default and does not mention an option to disable this feature.

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which discusses 'setting the maximum size of an attachment in XWikiPreferences', is ranked third. However, irrelevant nodes are ranked higher, such as the first node that 'does not address configuring mimetypes or setting maximum upload sizes', and the second node that 'does not provide information on configuring upload restrictions or size limits'.

ContextualRecall: The score is 0.64 because while the 3rd node in the retrieval context provides detailed steps for configuring the maximum attachment size (supporting sentences in Step 2 of the expected output), it lacks information on configuring allowed mimetypes in the xwiki.properties file (Step 1), which is why the score isn't higher.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific details on configuring 'xwiki.properties' and setting size to 10MB.

Correctness: Actual output specifies the maximum upload size correctly but configures mimetypes incorrectly by including PostScript and specific image formats instead of a wildcard. It also suggests irrelevant steps like disabling versioning.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.6254551924757624

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Correctness: 0.052731154854575066

Reasons:

AnswerRelevancy: The score is 0.80 because the response mostly addresses how to deny script rights to a space administrator, but includes an irrelevant statement about nested pages migration, which slightly detracts from its focus.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'programming rights and script execution' but doesn't address denying script rights to a space administrator. The second node provides a 'table of rights and their properties' without mentioning the denial of script rights. The third node explains 'script right and its default settings' but lacks information on denying these rights. The fourth node contains 'code snippets related to access checks' without addressing the denial of script rights. Lastly, the fifth node discusses

'converting rights and excluding pages' but is unrelated to denying script rights to a space administrator.
ContextualRecall: The score is 1.00 because the expected output is fully supported by the node in the retrieval context. Great job on achieving complete alignment!

CustomContextualRelevancy: The retrieval context confirms that rights implied by admin cannot be denied, aligning with the expected output. However, it does not explicitly state that XWiki does not support denying such rights.

Correctness: The actual output incorrectly states that script rights can be denied to a space administrator, which contradicts the expected output stating it is not supported in XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7342815424108319

Individual_scores:

AnswerRelevancy: 0.7647058823529411

Faithfulness: 0.7142857142857143

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.875

CustomContextualRelevancy: 0.6

Correctness: 0.785030991159669

Reasons:

AnswerRelevancy: The score is 0.76 because the output partially addresses the configuration question by providing relevant information, but it includes multiple irrelevant source links that do not directly contribute to solving the problem.

Faithfulness: The score is 0.71 because the actual output incorrectly claims that 'View' rights can be added at the wiki level for all teams, while the retrieval context clarifies that such permissions at the wiki-wide level are overridden by those set at a page level.

ContextualPrecision: The score is 0.67 because relevant nodes in the retrieval context are ranked higher than some irrelevant nodes, but not all. The first node discusses 'rights inheritance and administration features', which is relevant and ranked appropriately. However, the second node, which discusses 'parent/child relationship and nested spaces', is not directly related to configuring view and edit rights and should be ranked lower. Similarly, the third node about 'string formatting and user identification in OpenID Connect' is unrelated and should also be ranked lower. The fourth node correctly addresses the 'permission system in XWiki', making it relevant and well-placed. The fifth node, discussing 'creation of spaces and pages', is not directly related to the input and should be ranked lower. The sixth node provides information on 'setting wiki-wide and page-level rights', which is relevant and appropriately ranked.

ContextualRecall: The score is 0.88 because most of the expected output sentences are well-supported by the nodes in the retrieval context, such as the creation of groups and setting of rights at various levels. However, the lack of specific mention regarding the adjustment of rights for additional teams or spaces slightly reduces the score.

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and inheritance, aligning with the expected answer's steps on setting wiki-wide and space-level rights and inheriting rights for child pages. However, it lacks specific details on creating groups for each team and setting 'Edit' rights specifically for team groups, which are crucial steps in the expected answer.

Correctness: The actual output aligns with the expected output in setting view and edit rights, but lacks mention of 'Rights: Page & Children' for inheritance and refers to external links not present in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.4157461215034946

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 0.8

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.18165621620045497

Reasons:

AnswerRelevancy: The score is 0.85 because the response is mostly relevant to the issue of the unavailable bell icon, but it includes a source link that doesn't directly address the problem, slightly reducing its relevance.

Faithfulness: The score is 0.80 because the actual output incorrectly states that the Bell-Button is unavailable after installing the 'Notifications Application' extension, contradicting the retrieval context which indicates that this application adds a bell icon to the Alert menu in the top bar.

ContextualPrecision: The score is 0.17 because the relevant node in retrieval context, ranked sixth, discusses 'editing the xwiki.properties file', which is crucial for 'enabling notifications by setting `notifications.enabled` to `true`'. However, it is ranked lower than five irrelevant nodes. For instance, the first node discusses 'custom notification filters and their behavior', which is unrelated to the input about enabling notifications in the wiki settings.

ContextualRecall: The score is 0.50 because while the retrieval context mentions editing the xwiki.properties file, it does not explicitly state setting 'notifications.enabled' to 'true', which is a key part of the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications through the `xwiki.properties` file or setting `notifications.enabled` to `true`.

Correctness: The actual output does not mention setting `notifications.enabled` to `true` in `xwiki.properties` as required, and instead provides troubleshooting steps unrelated to the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.33375489015304133

Individual_scores:

AnswerRelevancy: 0.9375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.06502934091824791

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant to the issue of not receiving notifications, but it includes a source link that is not directly related to the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node contains 'image links and references that do not relate to notifications or their settings', the second node also includes 'image links and references unrelated to the topic of enabling notifications', and the third node similarly has 'image links and references that are not relevant to the issue of receiving notifications'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to the nodes in the retrieval context, which only contain image links and references unrelated to the content about notifications and following users or pages.

CustomContextualRelevancy: The retrieval context does not contain any information related to enabling notifications or following pages/users as described in the expected output.

Correctness: The actual output discusses issues with skins and system errors, which are unrelated to the expected steps for enabling notifications by following pages or users and setting configurations.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.31962421754569414

Individual_scores:

AnswerRelevancy: 0.6923076923076923

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.22543761296647252

Reasons:

AnswerRelevancy: The score is 0.69 because the response partially addresses the question about enabling notifications for one's own actions, but includes multiple irrelevant statements such as source links and apologies, which do not directly contribute to solving the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customization of page templates, internationalization, and other features unrelated to notification settings', the second node is about 'administrating wiki features and does not mention notification settings or user activity', and the third node covers 'various applications and extensions like Office Importer and Syntax Highlighting, which are not related to notification settings'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters as mentioned in the expected output.

Correctness: The actual output does not mention the key fact about disabling the filter in 'Erweiterten Filteroptionen,' and it includes speculative steps not present in the expected output. However, it slightly aligns by discussing notification settings.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.29113730928246584

Individual_scores:

AnswerRelevancy: 0.7222222222222222

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.024601633472572618

Reasons:

AnswerRelevancy: The score is 0.72 because the response addresses the main concern of disabling notifications effectively, but the repeated inclusion of unrelated source links slightly detracts from its overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Keep up the great work!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files', the second node is about 'moving attachments', the third node lists 'various macros', the fourth node covers 'administrating wiki features', and the fifth node talks about 'extensions and applications'. None of these relate to notification settings or how to manage them, which is the focus of the input.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to automatic page following or notification settings mentioned in the expected output.

Correctness: The actual output provides instructions on disabling notifications and references sources unrelated to the expected content about automatic tracking and notification settings.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.4445777807793698

Individual_scores:

AnswerRelevancy: 0.8125

Faithfulness: 0.875

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.2

Correctness: 0.1799666846762189

Reasons:

AnswerRelevancy: The score is 0.81 because the output mostly addresses the question about configuring XWiki for automatic user activation with OpenID Connect, but contains some irrelevant statements that slightly detract from its overall relevance.

Faithfulness: The score is 0.88 because the actual output incorrectly states that setting `oidc.skipped` to `false` ignores OpenID Connect, while the retrieval context indicates that OpenID Connect can be disabled using the property `oidc.skipped`.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'domain-based authentication and session cookies', which does not relate to 'automatic activation of user accounts upon first login with OpenID Connect'. The second node provides a 'general description of OpenID Connect authentication' without mentioning the specific property `oidc.enableUser` or its configuration in `xwiki.properties`. The third node includes 'metadata and compatibility information about the OpenID Connect Authenticator' but does not address the needed configuration. Lastly, the fourth node mentions the `xwiki.properties` file but lacks reference to

the `oidc.enableUser` property or the necessary configuration for automatic user activation.

ContextualRecall: The score is 0.60 because while the 4th node in retrieval context supports the mention of 'xwiki.properties' and 'oidc.enableUser' for automatic user account activation, it lacks specific instructions on updating the property, saving the file, and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions xwiki.properties but lacks details on enabling oidc.enableUser for auto-activation of accounts.

Correctness: The actual output mentions modifying `xwiki.cfg` and `oidc.skipped` which are not in the expected output. It misses adjusting `xwiki.properties` and `oidc.enableUser`. Context of automatic activation is maintained but with incorrect steps.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.4496234650880025

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.2977407905280153

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'the authentication management class and OpenID Connect endpoints,' but lacks information on 'user profile field customization or mapping in `xwiki.properties`.' The second node talks about 'different authentication methods like Container Authentication and Active Directory,' without addressing 'OpenID Connect user profile field customization.' The third node provides a 'general description of the OpenID Connect Authenticator and its installation,' but does not mention 'mapping user profile fields in `xwiki.properties`.' The fourth node gives 'metadata about the OpenID Connect Authenticator extension,' without discussing 'user profile field mapping.' Lastly, the fifth node lists 'dependencies for the OpenID Connect Authenticator extension,' which is unrelated to 'customizing user profile fields in `xwiki.properties`.'

ContextualRecall: The score is 0.40 because while the nodes in the retrieval context mention configuring user profile fields using `xwiki.properties` and mapping properties like `oidc.user.mapping`, they lack specific examples of custom property mapping, instructions on replacing placeholders, and details about saving and restarting XWiki.

CustomContextualRelevancy: The retrieval context lacks any information about customizing user profile fields using `oidc.user.mapping` in `xwiki.properties` as described in the expected output.

Correctness: The actual output does not mention configuring the `xwiki.properties` file or using `oidc.user.mapping`. It incorrectly suggests using `xwiki.cfg` and `oidc.userinfo.mappedfields`, which are not present in the expected output. The logical conclusion about mapping fields is present but lacks the specific method outlined in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.3697960575166574

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.21877634509994442

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about setting up group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node focuses on 'authentication management and OpenID Connect provider configuration, but does not mention group synchronization'. The second node discusses 'different authentication methods and extensions, but does not provide information on group synchronization'. The third node provides 'links to configuration guides and examples for OpenID Connect authentication but lacks details on configuring group synchronization'. The fourth node lists 'dependencies for the OpenID Connect authenticator extension, which is unrelated to the process of setting up group synchronization'. Finally, the fifth node describes 'the OpenID Connect authenticator's purpose and installation details, but does not address group synchronization setup'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks details about configuring group synchronization in xwiki.properties, such as setting oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping.

Correctness: Actual output mentions 'xwiki.cfg' instead of 'xwiki.properties', does not guide to set 'oidc.groups.claim', 'oidc.groups.separator', or 'oidc.groups.mapping'. Lacks specific setup details present in expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.5612135059833668

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0.325

ContextualRecall: 0.4

CustomContextualRelevancy: 0.3

Correctness: 0.5422810359002015

Reasons:

AnswerRelevancy: The score is 0.80 because the main content of the output is relevant to the input question about upload restrictions, but the repeated source link statements do not directly address the question, slightly reducing the relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the first three nodes in the retrieval context are irrelevant, discussing topics like 'trusted domains and PDF export customization', 'ODT export and PDF export features', and 'various features and macros', none of which relate to configuring attachment restrictions. Relevant nodes, which address 'Size Limit' and 'Mimetype Restriction', and provide 'detailed steps on setting the maximum size for attachments', are ranked lower. The presence of these relevant nodes prevents the score from being lower, but their lower ranking compared to irrelevant nodes keeps the score from being higher.

ContextualRecall: The score is 0.40 because while the retrieval context nodes provide detailed instructions for editing XWikiPreferences (related to sentences in Step 2 of the expected output), they lack information on configuring MIME types and setting maximum attachment sizes (related to sentences in Step 1 and the introductory sentence of Step 2).

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and MIME type restrictions but lacks specific details about configuring xwiki.properties for MIME types and setting size to 10 MB.

Correctness: The actual output suggests using the Attachment Validation Application instead of modifying xwiki.properties for MIME types, and it provides a less specific method for setting max upload size. It omits some steps and details found in the expected output, such as editing the XWikiPreferences object directly.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.33074363607496426

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2152310472190165

Reasons:

AnswerRelevancy: The score is 0.77 because the response partially addresses the input question, but includes irrelevant statements such as source links and an expression of inability to help, which do not directly answer the question about denying script rights to a space administrator.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in retrieval context are irrelevant. The first node discusses 'form validation and JavaScript patterns', the second node is about 'template providers and page creation', and the third node covers 'configuration classes and custom rights checks', none of which relate to denying script rights to an admin in XWiki.

ContextualRecall: The score is 0.00 because the expected output discusses rights and permissions in XWiki, which are not mentioned in any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about XWiki's admin rights or their limitations, which is the main focus of the expected output.

Correctness: The actual output provides contradictory information by suggesting steps to restrict rights, which the expected output claims cannot be denied. It also includes extraneous details about role creation and configuration, not present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.34137402183530247

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.29824413101181463

Reasons:

AnswerRelevancy: The score is 0.75 because the response partially addresses the question about configuring team access and editing permissions in XWiki, but includes several irrelevant links that do not directly relate to the specific task of setting permissions.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses XWiki's database support and programming features, which are not relevant to configuring access controls for teams. The second node is about the Notifications Application and its installation, which does not relate to configuring team access in XWiki. The third node explains Nested Pages and subwikis, which are not directly related to setting access permissions for different teams. The fourth node mentions managing users and groups, but does not provide specific instructions on configuring access controls for teams. The fifth node discusses group synchronization and customization, which is not directly relevant to the specific steps needed for configuring team access controls in XWiki.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output could be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to the steps for setting access control in XWiki as described in the expected output.

Correctness: The actual output suggests creating a separate Wiki for each team, which differs from creating groups within a single Wiki as in the expected output. It does not mention setting 'View' rights for 'XWikiAllGroup' and lacks detail on inheriting rights for child pages.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.4103031158142258

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.16181869488535477

Reasons:

AnswerRelevancy: The score is 0.80 because the response mostly addresses the issue of the notification bell not being available, but includes irrelevant links that do not directly pertain to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'custom notification filters and their behavior', the second node lists 'various links related to notification settings and customization', and the third node describes 'various notification preferences and settings', but none address 'enabling notifications in the wiki settings'. Similarly, the fourth node provides a 'description of the notification application and its features', and the fifth node discusses 'getting a notification RSS feed and notification settings', but neither mention 'enabling notifications in the wiki settings'.

ContextualRecall: The score is 0.50 because while the general idea of enabling notifications is supported by node 5 in the retrieval context, the specific instruction about setting `notifications.enabled` in `xwiki.properties` is not found in any node.

CustomContextualRelevancy: The retrieval context lacks any mention of enabling notifications via the `xwiki.properties` file, which is the core information in the expected output.

Correctness: The actual output mentions using a macro for notifications and references documentation, but it does not address enabling notifications via `xwiki.properties`, missing the key fact from the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.5187105378797149

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Correctness: 0.216109381124443

Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses the issue of not receiving notifications, but includes irrelevant links that do not directly help solve the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node 'primarily lists features and links related to the Notifications Application' but does not offer guidance on enabling notifications. The second node 'describes various settings and filters for notifications' without addressing the specific need to enable notifications. The third node talks about 'getting an RSS feed and using the Notifications Macro' but lacks instructions on enabling notifications. Lastly, the fourth node 'lists various notification preferences macros' without providing actionable steps on enabling notifications.

ContextualRecall: The score is 0.75 because while most aspects of enabling notifications and following users or pages are supported by nodes in the retrieval context, specific details like setting 'notifications.enabled' in 'xwiki.properties' and the 'réseau' tab for viewing followed users are not mentioned.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' to true and specific steps for following pages or users.

Correctness: The actual output does not match key facts such as enabling notifications via xwiki.properties or following users/pages. It provides different steps and lacks detailed guidance on following users or pages.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8360512881714092
Individual_scores:
AnswerRelevancy: 0.7777777777777778
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 1.0
CustomContextualRelevancy: 0.9
Correctness: 0.33852995125067714
Reasons:

AnswerRelevancy: The score is 0.78 because the response mostly addresses the question about enabling notifications for one's own actions, but it includes irrelevant links that do not directly help solve the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the first node in the retrieval context directly addresses the input question with 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' while the subsequent nodes discuss unrelated topics like 'RSS feeds and general notification settings,' 'links and headings related to notifications,' 'various notification preferences and macros,' and 'technical aspects of notifications,' which do not specifically address the issue of receiving notifications for one's own actions. Great job on ranking the relevant information first!

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the nodes in the retrieval context, showing a perfect match. Great job!

CustomContextualRelevancy: The retrieval context confirms that by default, users do not receive notifications for their own actions and mentions the ability to change this setting, aligning closely with the expected output. However, it does not explicitly mention the 'Options de filtrage avancées' section.

Correctness: The actual output provides steps to enable notifications for one's own actions, aligning partially with the expected output's mention of notifications for personal activities. However, it lacks the detail about the default settings and the specific 'Options de filtrage avancées' section.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7830411865004047
Individual_scores:
AnswerRelevancy: 0.8
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 1.0
CustomContextualRelevancy: 0.8
Correctness: 0.09824711900242791
Reasons:

AnswerRelevancy: The score is 0.80 because the response mostly addresses how to disable notifications, but it includes irrelevant links that do not directly help with the question, preventing a

higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are perfectly ranked higher than the irrelevant node. The first three nodes effectively address 'Choosing the default auto-watched mode', 'Notifications Auto Watch Preferences', and 'Auto Watch', which are directly related to the input query about modifying notification settings. The irrelevant node, which discusses 'RSS feeds and general notification settings', is appropriately ranked last.

ContextualRecall: The score is 1.00 because all sentences in the expected output are perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context covers auto-watched pages and custom notification filters, aligning with the expected answer about automatic page following and filter management.

However, it lacks explicit mention of modifying auto-watch settings in user profile notification settings.

Correctness: The Actual Output focuses on disabling unwanted notifications, which is unrelated to the Expected Output's topic of automatic page following and notification settings. No key facts align with the Expected Output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.44525977142872764

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.0

Correctness: 0.1429872000009374

Reasons:

AnswerRelevancy: The score is 0.93 because the output is highly relevant to configuring OpenID Connect for automatic user activation in XWiki, but it includes an irrelevant link to the LDAP Authenticator, which slightly detracts from the focus.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based authentication and session cookie configuration', which does not relate to enabling user accounts on first login using OpenID Connect. The second node mentions 'container authentication and OpenID Connect', but lacks information on automatic user account activation. The third node focuses on 'authentication and synchronization', not on enabling user accounts automatically. The fourth node lists 'dependencies for the OpenID Connect extension', which is not relevant to the required configuration. Lastly, the fifth node is about 'LDAP and Active Directory integration', unrelated to the OpenID Connect setup needed.

ContextualRecall: The score is 0.60 because while the retrieval context nodes mention 'OpenID Connect Authentication' and 'xwiki.properties', which supports the general topic of adjusting the file for OpenID Connect, they do not provide specifics on adding or updating properties, saving the file, or restarting XWiki as mentioned in the expected output.

CustomContextualRelevancy: The retrieval context does not mention adjusting the xwiki.properties file or the specific property oidc.enableUser for automatic user account activation with OpenID Connect in XWiki.

Correctness: The actual output refers to `oidc.createuser` in `xwiki.cfg` instead of `oidc.enableUser` in `xwiki.properties`. It discusses account creation, not activation, and uses a different configuration file and property name than the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.6025079299332448

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.32615869071057957

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant and provides useful information, but it includes sources and a documentation link that do not directly address the specific question about customizing user profile fields in XWiki using OpenID Connect.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions the 'xwiki.properties' file and its relevance for mapping user profile fields, is ranked first. Irrelevant nodes, such as those discussing container authentication and LDAP (second node), general information about the OpenID Connect Authenticator extension (third node), dependencies for the OpenID Connect extension (fourth node), and configuration guides and limitations without discussing 'xwiki.properties' (fifth node), are correctly ranked lower. Great job on getting it spot on!

ContextualRecall: The score is 0.40 because while nodes in the retrieval context mention customizing user profile fields and using 'oidc.user.mapping', they lack specific examples and details about replacing property names, using variables, and the process of saving and restarting XWiki.

CustomContextualRelevancy: The retrieval context does not mention customizing user profile fields or using 'oidc.user.mapping' in 'xwiki.properties' as described in the expected output.

Correctness: Actual output uses 'xwiki.cfg' instead of 'xwiki.properties' for mapping. It introduces extraneous information like default fields filled by OpenID Connect and specific examples not present in expected output. Core concept of mapping is maintained, but discrepancies in file and property names reduce alignment.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6413311873798654

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.9166666666666666

CustomContextualRelevancy: 0.3

Correctness: 0.2979871242791926

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider, with no irrelevant statements.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which is the third node, 'explicitly mentions group synchronization, the use of claims in the xwiki.properties file, and the oidc.groups.mapping property, which are relevant to the expected output.' However, it is ranked lower than the first and second nodes, which 'discuss authentication management and endpoints related to OpenID Connect, but do not address group synchronization specifically,' and 'mention various authentication methods and extensions, but do not provide information on configuring group synchronization with OpenID Connect,' respectively. This lowers the precision score as irrelevant nodes are ranked higher than the relevant one.

ContextualRecall: The score is 0.92 because most instructions in the expected output, such as configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', are supported by the 3rd node in the retrieval context. However, the retrieval context does not explicitly mention saving the xwiki.properties file and restarting XWiki, which affects the score.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using oidc.groups.mapping but lacks specific details about oidc.groups.claim and oidc.groups.separator properties, which are crucial in the expected output.

Correctness: The actual output lacks specific instructions on setting 'oidc.groups.claim' and 'oidc.groups.separator' found in the expected output. It introduces new concepts like 'oidc.userinfoclaims' and listeners which are not mentioned in the expected output. The mapping of groups is addressed but with different format details.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.2700045495034044

Individual_scores:

AnswerRelevancy: 0.5833333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.03669396368709306

Reasons:

AnswerRelevancy: The score is 0.58 because the answer includes multiple irrelevant source links that do not directly address the question about setting file upload restrictions, which lowers the relevancy. However, the score is not lower because there is some relevant information present that partially addresses the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'language preferences and translations in XWiki', which is unrelated to configuring file upload restrictions. The second node lists 'various XWiki applications and features, such as tour guides and mail configuration', but does not mention file upload settings. The third node explains 'PDF export customization using CSS and XSL', which is not relevant to setting file upload restrictions. The fourth node provides 'information about system notification filters and office document

viewing', which are unrelated to file upload configurations. The fifth node contains 'image links and references to various projects', which do not pertain to configuring file upload restrictions. Lastly, the sixth node includes 'image links and references to various organizations', which are not relevant to the topic of file upload settings.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be traced back to any nodes in the retrieval context, as the context lacks information on file size limits, attachment restrictions, MIME type configurations, and instructions for editing or saving changes in `XWikiPreferences`.

CustomContextualRelevancy: The retrieval context does not contain any information related to configuring MIME types or attachment size limits in XWiki, which are key aspects of the expected output.

Correctness: The actual output does not provide any specific steps or facts about MIME type configuration or setting file size limits to 10 MB, as outlined in the expected output. It only offers general suggestions without concrete actions.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.6082201174753871

Individual_scores:

AnswerRelevancy: 0.7272727272727273

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.13871464424626162

Reasons:

AnswerRelevancy: The score is 0.73 because the response addresses the question about refusing script rights to a space administrator, but it includes irrelevant source links that do not directly relate to the query, slightly detracting from its overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 0.58 because the first node in the retrieval context, which is about 'language preferences and translations in XWiki,' is irrelevant to script rights and administrative permissions, yet it is ranked higher than relevant nodes. Additionally, the fourth node, which discusses 'checking access rights programmatically,' and the fifth node, focusing on 'executing code with specific user rights,' are also irrelevant to the input question and are ranked higher than they should be. However, the second and third nodes are relevant as they discuss 'programming and script rights' and specify that 'ADMIN rights are not deniable,' which supports the input query.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the retrieval context, specifically by the 3rd node, which confirms that admin rights are not deniable. Great job!

CustomContextualRelevancy: The retrieval context mentions that ADMIN rights imply other rights and are not deniable, but does not explicitly state that administrator rights cannot be refused as in the expected output.

Correctness: The actual output contradicts the expected output by stating that script rights can be denied to a space admin, whereas the expected output states admin rights cannot be refused.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.36714377029818307

Individual_scores:

AnswerRelevancy: 0.8518518518518519

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.3510107699372468

Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses the question about configuring spaces in XWiki, but it includes multiple irrelevant statements providing source links that do not directly answer the question.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses the structure of an XML page in XWiki, which does not relate to configuring access rights for teams. The second node talks about sending representations and modifying resources, which is not directly related to configuring team access rights. The third node is about managing comments and attachments, which is not relevant to setting up access rights for teams. The fourth node explains spaces and sub-pages but does not provide information on configuring access rights. The fifth node lists actions for running a migrator and handling space names, which is unrelated to configuring team access rights. Finally, the sixth node discusses configuration display scopes, which does not address setting up access rights for teams.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to creating groups, defining rights, or inheriting rights as outlined in the expected output.

Correctness: The actual output discusses creating user groups and setting permissions but lacks clarity and detail on setting 'Afficher' rights for 'XWikiAllGroup' and inheriting rights for child pages, which are key points in the expected output. It also introduces unnecessary technical details not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.6978200550989072

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Correctness: 0.505967949641062

Reasons:

AnswerRelevancy: The score is 0.67 because the response addresses the question about authenticating users with access tokens, but it includes multiple irrelevant source links that do not directly contribute to the answer.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because relevant nodes in the retrieval context are mixed with irrelevant ones. For example, the first node discusses 'LDAP configuration, which is not relevant to token-based authentication,' yet it is ranked higher than the second node, which 'explains how to configure authorized applications, relevant to using access tokens for authentication.' Similarly, the third node about 'OpenID Connect scopes and response types' is not directly related to the input and is ranked higher than the fourth node, which 'describes the use of JSON Web Tokens for authentication,' directly relating to token-based authentication. This mix of relevant and irrelevant nodes affects the precision score.

ContextualRecall: The score is 0.71 because while several key aspects of the expected output, such as the use of JWT for authentication and the configuration of authorized applications, are supported by nodes in the retrieval context, specific details like generating a JWT token with required claims and including it in the 'Authorization' header are not explicitly mentioned in any node.

CustomContextualRelevancy: The retrieval context confirms the use of JWT for authentication and mentions configuring authorized applications, matching the expected answer. However, it lacks specific details on JWT claims and token inclusion in the request header.

Correctness: The actual output correctly mentions configuring the `xwiki.authentication.authclass` property for token-based authentication and using JWT, but it omits details about the required claims in the token and Ed25519 signing key. It also lacks information on configuring authorized applications and the specific use of the 'Authorization' header.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.31927358701918823

Individual_scores:

AnswerRelevancy: 0.7272727272727273

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.18836879484240196

Reasons:

AnswerRelevancy: The score is 0.73 because the answer mostly addresses the question about authenticating users with access tokens, but it includes irrelevant source links that do not directly contribute to the explanation.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'model validation strategies and configuration in XWiki', which is unrelated to user authentication with access tokens. The second node is about 'attachment validation and configuration, including mimetype restrictions', which does not relate to authenticating users with access tokens. The third node provides information about 'the admin UI for model validation and its installation', which is not relevant to the topic of user authentication using access tokens.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, as the context lacks any mention of key concepts such as 'Benutzer mit Zugriffstoken', 'LLM-Anwendung (BETA)-Erweiterung', 'externen Anwendungen', 'signierten JSON Web Tokens (JWT)', 'Token-Authentifizierer', 'xwiki.authentication.authclass', 'autorisierten Anwendungen', 'JWT-Token', 'Authorization-Header', or 'Erstellung oder Aktualisierung von Benutzern'.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are key components of the expected output.
Correctness: The actual output mentions authentication with tokens but lacks detail on JWT, Ed25519 signing, and specific configuration steps outlined in the expected output.
Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_033_result.json

Overall_score: 0.2502112251030758

Individual_scores:

AnswerRelevancy: 0.3

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.20126735061845485

Reasons:

AnswerRelevancy: The score is 0.30 because the response includes multiple irrelevant links and statements that do not directly address the question about authenticating users with access tokens. However, there might still be some relevant information present, which is why the score is not lower.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node is about 'upgrading XWiki' and does not mention 'user authentication with access tokens'. The second node describes 'features of the Administration Application in XWiki', which also does not address 'authentication with access tokens'. The third node discusses 'managing user rights and profiles in XWiki', but again does not provide information on 'authenticating users with access tokens'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information about token-based authentication, JWT, or related configuration details mentioned in the expected output.

Correctness: The actual output lacks details on JWT configuration such as specific claims, Ed25519 key, and the steps to enable JWT authentication outlined in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_mistral-nemo_12b_Q4

File: qa_001_result.json

Overall_score: 0.2779194661310235

Individual_scores:

AnswerRelevancy: 0.9259259259259259

Faithfulness: 0.6666666666666666

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.0749242041935482

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and addresses the issue with the bell icon in XWiki. However, it includes a source link that does not directly contribute to solving the problem, preventing a perfect score.

Faithfulness: The score is 0.67 because the actual output incorrectly states that the bell icon for notifications appears only if there is at least one unread notification, while the retrieval context does not specify this condition.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question. The first node discusses 'installation instructions and prerequisites for extensions,' which does not help with enabling notifications. The second node describes 'the functionality of the Alert menu,' but lacks information on enabling notifications. The third node lists 'dependencies related to the notifications extension,' yet fails to address enabling notifications. Lastly, the fourth node talks about 'overriding notification templates,' which is unrelated to the issue of enabling notifications.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications through the `xwiki.properties` file or the `notifications.enabled` setting.

Correctness: The actual output does not address enabling notifications via the `notifications.enabled` setting in `xwiki.properties`, which is crucial in the expected output. It instead focuses on troubleshooting the bell icon and dependencies.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6566651881259623

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8

ContextualPrecision: 0.8055555555555555

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Correctness: 0.33443557320021833

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the issue of not receiving notifications in XWiki. Great job!

Faithfulness: The score is 0.80 because the actual output incorrectly states that no application types are enabled for notifications by default in XWiki 15.5 and later, except specific ones like Mention, whereas the retrieval context clarifies that all application types are enabled by default. Additionally, the actual output inaccurately claims you can watch a user from their profile page in the 'network' tab, while the retrieval context specifies that the 'network' tab only shows the list of users you are watching, not a way to watch a user from there.

ContextualPrecision: The score is 0.81 because most relevant nodes in the retrieval context are ranked higher than irrelevant ones. For example, the first node explains 'how to follow a user to receive notifications', aligning well with the input query. However, the second node, which 'discusses filtering notifications and hidden pages', is not directly related to enabling notifications and should be ranked lower than the third node, which 'mentions the settings for notifications'. Similarly, the fifth node introduces 'the 'What's New' application', which is not directly related to the issue of not receiving notifications and should be ranked lower than the fourth node, which 'provides an overview of different ways to receive notifications in XWiki'.

ContextualRecall: The score is 0.50 because while some details like using the 'alert' menu and following users for notifications are supported by nodes in the retrieval context, key instructions such as enabling notifications via `xwiki.properties` and subscribing to pages or wikis are not covered.

CustomContextualRelevancy: The retrieval context includes information on following users and watching pages, but lacks details on enabling notifications via 'xwiki.properties' and the alert menu specifics.

Correctness: The actual output mentions that notifications are not enabled by default, which is not stated in the expected output. It correctly identifies the need to watch users or spaces and discusses notification settings, but adds information about filters and default settings that are not present in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9790043232107569

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.8740259392645413

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about enabling notifications for one's own activity in XWiki. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than irrelevant nodes. The first node explains, 'By default, one don't receive notification about one's own activity... You have the ability to disable this filter in the "Advanced filtering options" section of your own notification settings,' which directly answers the input question. The second node also supports this by stating, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' which is relevant to enabling notifications for one's own actions. The irrelevant nodes, such as the third node discussing 'administrator settings and notification preferences inheritance,' do not directly address the input question and are ranked lower. Great job maintaining perfect precision!

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context fully contains the expected answer, detailing the default notification settings and the ability to change them in the 'Advanced filtering options' section.

Correctness: The actual output accurately reflects the expected output with minor discrepancies; it adds steps for disabling the filter, aligning with the expected logical flow and facts.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5924714086474765

Individual_scores:

AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 1.0
CustomContextualRelevancy: 0.3
Correctness: 0.2548284518848593

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about disabling notifications in XWiki without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'how to get a notification RSS feed and general notification settings', which does not address disabling notifications or automatic page watching. The second node is about 'following a user and default watch settings', which also does not mention disabling notifications or adjusting automatic page watching settings. Subsequent nodes similarly fail to address the specific issue of disabling notifications or automatic page watching, focusing instead on unrelated aspects like 'filters for notifications', 'enabling or disabling email notifications', 'administrator settings', and 'overriding default notification templates'.

ContextualRecall: The score is 1.00 because all sentences in the expected output are fully supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context mentions default watch settings and user profile settings but lacks details on automatic page watching and deleting notification filters.

Correctness: The actual output discusses managing notification settings but lacks specifics on 'automatic page watching' and 'deleting custom filters' from the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8620039661838952
Individual_scores:
AnswerRelevancy: 0.9090909090909091
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 0.4
CustomContextualRelevancy: 1.0
Correctness: 0.8629328880124627

Reasons:

AnswerRelevancy: The score is 0.91 because the response is highly relevant and mostly addresses the question about configuring OpenID Connect in XWiki. However, it includes an irrelevant statement about the LDAP Authenticator, which slightly detracts from its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job maintaining accuracy!

ContextualPrecision: The score is 1.00 because the relevant node, which mentions the property '#oidc.enableUser=true' directly related to enabling user accounts on first login, is ranked first. This ensures that the most pertinent information is prioritized for the input query. Great job on achieving perfect precision!

ContextualRecall: The score is 0.40 because while the retrieval context (node 1) supports the mention of enabling user accounts on first login and the specific property 'oidc.enableUser=true', it lacks information on adding or updating properties, saving the file, restarting XWiki, and ensuring automatic

account enablement.

CustomContextualRelevancy: The retrieval context contains the exact information needed to enable user accounts on first login by setting 'oidc.enableUser=true' in the xwiki.properties file, matching the expected output.

Correctness: The actual output correctly describes modifying the 'oidc.enableUser' property in 'xwiki.properties' to enable accounts on first login. However, it omits the step to uncomment the property, which is explicitly mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.8734007422660065

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Correctness: 0.6404044535960394

Reasons:

AnswerRelevancy: The score is 1.00 because the answer perfectly addresses the question without any irrelevant statements. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job maintaining accuracy!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context is ranked first, providing direct information on how to 'associate non standard properties coming from the OpenID Connect provider with the XWiki user', which is exactly what the input is asking about. The irrelevant nodes, which discuss topics such as 'access to various claims', 'configuration of OpenID Connect endpoints', 'group synchronization', and 'in-wiki configuration', are correctly ranked lower as they do not pertain to mapping user profile fields.

ContextualRecall: The score is 0.60 because while the retrieval context (node 3) supports the mapping example and customization process described in the expected output, it lacks explicit details on replacing 'myxproperty1' and 'myxproperty2' with actual XWiki user property names, and does not mention the need to save the 'xwiki.properties' file and restart XWiki.

CustomContextualRelevancy: The retrieval context contains all the information from the expected output, including mapping user profile fields using the 'oidc.user.mapping' property in the 'xwiki.properties' file and the example mappings provided.

Correctness: The actual output includes the necessary mapping example and explanation for 'oidc.user.mapping', aligning with the expected output. However, it introduces additional facts like 'oidc.user.nameFormatter', group synchronization, and event listeners, which are not in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8359659511269095

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.46153846153846156

CustomContextualRelevancy: 0.9

Correctness: 0.6542572452229954

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant ones. The first three nodes provide detailed information on configuring properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, which are crucial for group synchronization as required by the input. The irrelevant nodes, ranked fourth and fifth, discuss general OpenID Connect information and authentication management, which do not directly pertain to the group synchronization process.

ContextualRecall: The score is 0.46 because while some specific configuration details like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` are supported by nodes in the retrieval context, many introductory and procedural instructions such as locating and editing the `xwiki.properties` file, as well as saving changes and restarting XWiki, are missing from the retrieval context.

CustomContextualRelevancy: The retrieval context includes information about configuring group synchronization using the `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` properties, which matches the expected output. However, it lacks explicit instructions on locating and editing the `xwiki.properties` file and restarting the XWiki instance.

Correctness: The actual output correctly addresses group synchronization setup, claim configuration, and group mapping, similar to the expected output. However, it introduces `oidc.userinfoclaims` not present in the expected output and lacks the step to save and restart XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6078383899888478

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.6363636363636364

CustomContextualRelevancy: 0.2

Correctness: 0.5440000369027834

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to the question about configuring attachment settings, but it includes a link to a general guide on XWiki basic concepts, which is not directly relevant to the specific question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which 'discusses setting limits on the maximum size of an attachment and provides steps to change it', is ranked third. This node should be ranked higher than the first node, which 'discusses linking to external

files and images, which is unrelated to configuring upload restrictions or size limits', and the second node, which 'describes the structure of XWiki pages and their components, which does not relate to setting upload restrictions or size limits'.

ContextualRecall: The score is 0.64 because while several detailed instructions (sentences 2 to 9) are supported by nodes in the retrieval context, the introductory statements and explanations (sentences 1, 3, and 7) lack direct support from the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions, but lacks specific details on configuring 'xwiki.properties' for mimetypes and setting size to 10MB.

Correctness: The actual output correctly outlines the steps to set size limit and mimetype restriction but incorrectly includes 'attachment.download.forceDownload' and 'attachment.upload.enableComments', which are not part of the expected output. It also omits the detailed steps for configuring the maximum attachment size in the XWiki user interface.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.627108410070746

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.0483647461387619

Reasons:

AnswerRelevancy: The score is 0.71 because the answer partially addresses the question about denying script rights, but includes irrelevant source links and markdown format statements that do not directly contribute to the solution.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining accuracy and alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses programming rights and script execution but does not address the denial of script rights to a space administrator. The second node lists rights and their default states but does not mention the inability to deny rights implied by admin rights. The third node explains the 'Script' right but does not mention the inability to deny rights implied by admin rights. The fourth node provides technical details about checking access rights but does not address the specific issue of denying script rights to a space administrator. The fifth node discusses converting rights during page migration, which is unrelated to denying script rights to a space administrator.

ContextualRecall: The score is 1.00 because every part of the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context confirms that rights implied by admin cannot be denied, matching the expected output.

Correctness: The actual output incorrectly suggests that the 'Script' right can be denied to a space administrator, contradicting the expected output that states implied rights by admin cannot be denied.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.6854949036976433

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0.29166666666666663

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Correctness: 0.9531209373373757

Reasons:

AnswerRelevancy: The score is 0.82 because the output mostly addresses the configuration question in XWiki, but it includes multiple source links that do not directly contribute to solving the problem, slightly reducing the relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.29 because the relevant nodes in the retrieval context are ranked lower than irrelevant nodes. The first node discusses 'content organization and rights inheritance in XWiki,' which does not provide specific steps for configuring permissions for teams, making it less relevant. The second node explains 'the parent/child relationship and the introduction of Nested Pages in XWiki,' which also does not address team access rights. The third node is about 'string formatting and OpenID Connect properties,' unrelated to the query. These irrelevant nodes are ranked higher than the fourth node, which provides relevant information about 'XWiki's permission system, including setting view and edit rights.' Additionally, the sixth node, which outlines 'basic rules for setting rights in XWiki,' is also relevant but ranked lower. This misranking of nodes contributes to the lower score.

ContextualRecall: The score is 0.75 because while most steps in the expected output align well with the nodes in retrieval context, such as setting rights and inheriting them for child pages, specific actions like creating groups for each team and adjusting rights for additional teams or spaces are not directly supported by the nodes.

CustomContextualRelevancy: The retrieval context mentions rights inheritance and setting rights at different levels, but lacks specific details about creating groups for each team and setting rights specifically for 'XWikiAllGroup' or individual team spaces as described in the expected output.

Correctness: All factual statements in the actual output align with the expected output, including creating groups, assigning wiki-wide view rights, and setting space-specific edit rights with inheritance for child pages. No discrepancies or deviations are present.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.44757094163657607

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.2187589831527894

Reasons:

AnswerRelevancy: The score is 0.80 because the main issue about the notification bell not being available is addressed, but the inclusion of source links that do not directly address this issue prevents the score from being higher.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node, which discusses 'editing the xwiki.properties file' to enable notifications, is ranked last (6th). The irrelevant nodes, such as the first node discussing 'custom notification filters and their behavior', should be ranked lower as they do not address enabling notifications in the wiki settings.

ContextualRecall: The score is 0.50 because while the 5th node in the retrieval context supports the idea of editing the xwiki.properties file, it does not specifically mention setting 'notifications.enabled' to 'true', which is a key detail in the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via 'xwiki.properties' or setting 'notifications.enabled' to 'true'.

Correctness: The actual output suggests troubleshooting steps and provides links but does not mention setting 'notifications.enabled' in 'xwiki.properties' to 'true', which is a key detail in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.321351144063235

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.01144019771274345

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant to the issue of not receiving notifications, but it includes a source link that does not directly address the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job maintaining accuracy!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node contains 'image links and references that are unrelated to the topic of notifications or their settings.' Similarly, the second node is 'a list of image links and references, which do not provide any information about enabling notifications or related settings.' The third node also consists of 'image links and references, which are not relevant to the issue of receiving notifications or their configuration.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to the nodes in the retrieval context, which only contain image links and references.

CustomContextualRelevancy: The retrieval context contains image links and references unrelated to the expected answer about enabling notifications and following users or pages in XWiki.

Correctness: The actual output does not mention enabling notifications via 'xwiki.properties' or following pages/users as stated in the expected output, and includes irrelevant information about email settings and server issues.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.31178604733692056

Individual_scores:

AnswerRelevancy: 0.72727272727273

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.14344355674879614

Reasons:

AnswerRelevancy: The score is 0.73 because the response partially addresses the question about enabling notifications for one's own activities, but includes irrelevant source links and information about XWiki features that do not directly answer the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customization of page templates, internationalization, and other features unrelated to notification settings,' the second node is about 'administrating wiki features and does not mention notification settings or user activity,' and the third node focuses on 'importing documents, sharing pages, and other features not related to notifications for user activity.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, indicating a complete mismatch between the expected information and the provided context.

CustomContextualRelevancy: The retrieval context lacks any information related to notification settings or user activity, which are key elements in the expected output.

Correctness: The actual output does not address the expected output about disabling a notification filter in user settings and instead provides general instructions for a different system.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.2711374799528692

Individual_scores:

AnswerRelevancy: 0.5384615384615384

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.08836334125567688

Reasons:

AnswerRelevancy: The score is 0.54 because while the output attempts to address the question about disabling notifications, it includes several irrelevant links and explanations that do not directly help in resolving the issue, thus lowering the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files and various related features', which does not pertain to 'notification settings'. Similarly, the second node is about 'moving attachments', and the third node lists 'various macros and features in XWiki', neither of which are related to managing notifications. The fourth node focuses on 'administrating wiki features like users and rights', and the fifth node is about 'controlling permissions within the wiki', both unrelated to notification settings. Lastly, the sixth node lists 'various extensions like Office Importer and Syntax Highlighting', which are also irrelevant to the input query about disabling notifications.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to automatic page following, notification settings, or custom filters as mentioned in the expected output.

Correctness: The actual output does not match the expected output in any key facts. It discusses attachment management, not notification settings as in the expected output. It lacks critical information about automatic following of pages and modifying notification settings, and includes unrelated troubleshooting steps.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.46668695543533345

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.4001217326120007

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring XWiki for automatic user activation with OpenID Connect. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based authentication and session cookie configuration,' which does not relate to automatically activating user accounts with OpenID Connect. The second node provides 'general information about OpenID Connect authentication,' but lacks specifics on auto-activating user accounts. The third node includes 'metadata about the OpenID Connect Authenticator extension,' which is not pertinent to the configuration change needed. The fourth node mentions 'provider configuration and endpoints,' but not the specific property needed for auto-activation.

ContextualRecall: The score is 0.40 because while there is some alignment with nodes in the retrieval context regarding the file `xwiki.properties` and the property `oidc.enableUser`, key instructions like adding or updating the property, setting it to `true`, and the outcome of automatic account activation are missing.

CustomContextualRelevancy: The retrieval context does not contain any information about modifying the xwiki.properties file or the oidc.enableUser property to automatically activate user accounts on first login with OpenID Connect.

Correctness: Actual output uses 'xwiki.automaticUserActivation=true' instead of 'oidc.enableUser=true'. It adds unnecessary steps about installing an add-on and lacks details on uncommenting the property.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.2971154071196393

Individual_scores:

AnswerRelevancy: 0.5151515151515151

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.26754092756632053

Reasons:

AnswerRelevancy: The score is 0.52 because the output contains multiple irrelevant statements about importing Java classes and method definitions, which do not directly address the question about customizing user profile fields in XWiki. However, some relevant information may still be present, contributing to the score being above zero.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'general OpenID Connect configuration and endpoints' without mentioning user profile field mapping. The second node is about 'authentication methods and extensions, including LDAP and Active Directory,' which does not address OpenID Connect user profile field mapping. The third node provides 'links to configuration guides for specific systems' but lacks information on mapping user profile fields in `xwiki.properties`. The fourth node offers 'general information about the OpenID Connect Authenticator extension' without discussing user profile field mapping. Lastly, the fifth node lists 'dependencies for the OpenID Connect Authenticator extension,' unrelated to mapping user profile fields in `xwiki.properties`.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention customizing user profile fields or using `oidc.user.mapping` in `xwiki.properties`, which are key elements in the expected answer.

Correctness: The actual output incorrectly references 'xwiki.cfg' instead of 'xwiki.properties', omits the use of 'oidc.user.mapping', and introduces unnecessary scripting details not mentioned in the expected output. It also lacks the step to restart XWiki after saving changes.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.3743717193493307

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.07692307692307693

CustomContextualRelevancy: 0.0

Correctness: 0.31216438203005015

Reasons:

AnswerRelevancy: The score is 0.86 because the response largely addresses the process of setting up group synchronization between XWiki and an OpenID Connect provider, but it includes source links that do not directly contribute to explaining the setup process.

Faithfulness: The score is 1.00 because there are no contradictions, indicating the actual output aligns perfectly with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and endpoint configurations', which do not relate to group synchronization. The second node focuses on 'container authentication and other

authentication methods', missing the topic of group synchronization. The third node mentions 'authentication on an OpenID Connect provider' without detailing group synchronization. The fourth node lists 'dependencies for the OpenID Connect extension', which is unrelated to group synchronization. Lastly, the fifth node covers 'the OpenID Connect authenticator's features and installation details', but not the specifics of group synchronization.

ContextualRecall: The score is 0.08 because only the mention of 'xwiki.properties' in the 1st node in retrieval context supports the expected output, while all other instructions and configurations are unsupported by the nodes.

CustomContextualRelevancy: The retrieval context lacks specific details about configuring group synchronization, such as setting 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' in 'xwiki.properties', which are present in the expected output.

Correctness: The actual output mentions group synchronization but uses different file names (xwiki.cfg instead of xwiki.properties) and lacks details on group separator and mapping configurations present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.5432391980774556

Individual_scores:

AnswerRelevancy: 0.6153846153846154

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.63636363636364

CustomContextualRelevancy: 0.3

Correctness: 0.5076869367164816

Reasons:

AnswerRelevancy: The score is 0.62 because the output contains several irrelevant source links that do not directly address the question about setting upload restrictions for images or PDF files. However, the score is not lower because there is some relevant information that partially addresses the input question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.20 because only the fifth node in the retrieval context is relevant, as it 'includes detailed steps for setting the maximum attachment size and mentions mimetype restrictions, directly relevant to the expected output.' The first four nodes should be ranked lower because they discuss 'trusted domains and PDF export customization,' 'ODT and PDF export features,' 'a list of macros and features,' and 'size limits and mimetype restrictions' without providing specific instructions, which are not relevant to configuring attachment upload restrictions. Additionally, the sixth node is about 'document lifecycle features,' which does not address attachment upload restrictions.

ContextualRecall: The score is 0.64 because while the 5th node in the retrieval context covers many aspects of configuring maximum upload size and navigating XWikiPreferences (as seen in sentences 4-8 of the expected output), it lacks details on MIME type restrictions and specific file type limitations mentioned in sentences 1-3.

CustomContextualRelevancy: The retrieval context mentions setting limits on maximum attachment size and MIME type restrictions, but lacks specific details about configuring MIME types in xwiki.properties and setting size to 10 MB.

Correctness: The actual output captures the maximum upload size configuration and need to repeat steps for each subwiki, but misses configuring MIME types in 'xwiki.properties' and instead suggests using an extension. Terminology and sequence differ, impacting consistency.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_019_result.json

Overall_score: 0.3295073532087731
Individual_scores:
AnswerRelevancy: 0.8571428571428571
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.11990126210978155
Reasons:

AnswerRelevancy: The score is 0.86 because the response mostly addresses the input question about denying script rights to a domain administrator, but includes irrelevant source links that do not directly contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'JavaScript validation and field matching', which is unrelated to the topic of denying script rights in XWiki. The second node is about 'template providers and page creation in XWiki', which does not relate to script rights or their denial. The third node describes 'a custom right check configuration', but it does not address the specific issue of denying script rights to a space administrator in XWiki. Therefore, the irrelevant nodes are not ranked lower than any relevant nodes, resulting in a score of 0.00.

ContextualRecall: The score is 0.00 because the expected output sentence about a feature not supported by XWiki, specifically regarding admin rights, is not mentioned in any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention anything about admin rights or their implications in XWiki, which is the focus of the expected output.

Correctness: The actual output provides steps to modify permissions in XWiki, while the expected output states that denying rights implied by admin rights is not supported, showing a discrepancy in key facts.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_020_result.json

Overall_score: 0.4319277332831805
Individual_scores:
AnswerRelevancy: 0.8148148148148148
Faithfulness: 1.0
ContextualPrecision: 0.25
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.5267515848842679
Reasons:

AnswerRelevancy: The score is 0.81 because the output largely addresses the question about configuring permissions in XWiki, but includes several irrelevant links to unrelated documentation,

which slightly detracts from its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.25 because the fourth node in the retrieval context, which is relevant as it 'provides a link to managing users and groups,' is ranked lower than three irrelevant nodes. The first node discusses 'XWiki's use of databases and programming features,' the second node describes 'the Notifications Application,' and the third node explains 'Nested Pages and subwikis,' none of which are pertinent to configuring access controls for teams. This misranking of relevant and irrelevant nodes results in a lower score.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context. The retrieval context lacks information on steps for access control, creating groups, setting rights, and ensuring visibility and edit permissions as described in the expected output.

CustomContextualRelevancy: The retrieval context lacks any information related to access control steps or team-specific rights management as outlined in the expected output.

Correctness: The actual output suggests using nested pages and specific permissions but lacks detailed steps on setting wiki-wide and area-level rights, and inheritance of rights for subpages, as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_021_result.json

Overall_score: 0.3175517091784796

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.030310255070877527

Reasons:

AnswerRelevancy: The score is 0.88 because the response is mostly relevant to the issue of the bell icon not being available, but it includes a link to XWiki syntax documentation, which is not directly related to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'custom notification filters and their behavior', which is unrelated to the input about enabling notifications in the wiki settings. Similarly, the second node lists 'various notification settings and customization options' without addressing the enabling of notifications. The third node provides 'links to various notification preferences' but does not mention enabling them. The fourth node describes 'the notification menu and its features', which is not related to the input. Lastly, the fifth node discusses 'getting an RSS feed for notifications and notification settings', again not mentioning enabling notifications in the wiki settings.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, which lacks specific information about enabling notifications or modifying `xwiki.properties`.

CustomContextualRelevancy: The retrieval context does not mention setting `notifications.enabled` in `xwiki.properties` to `true`, which is the core information in the expected output.

Correctness: The actual output does not address enabling notifications via the `xwiki.properties` file, and introduces unrelated information about a bell icon and potential translation errors, which are not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_022_result.json

Overall_score: 0.6778573772244777

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.4

Correctness: 0.27240742124160283

Reasons:

AnswerRelevancy: The score is 0.89 because the response mostly addresses the issue of not receiving notifications, but includes irrelevant source links that do not directly relate to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes are perfectly ranked higher than irrelevant ones. The first two nodes are relevant as they discuss 'Enable and disable the extension' and 'Enabling notifications for all users by default', and 'Follow a user' and 'Watched Entities', which align with the input about notifications. The irrelevant nodes, ranked third and fourth, discuss 'RSS feeds and settings' and 'various notification preferences macros', which do not directly relate to enabling notifications or following users/pages.

ContextualRecall: The score is 0.50 because while there are nodes in the retrieval context that support the general idea of enabling notifications and following users or pages (1st and 2nd nodes), there are several specific details in the expected output, such as 'xwiki.properties', 'menu alerte', and 'réseau' tab, that are not covered by the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting `notifications.enabled` to `true` and specifics about the alert menu or network tab in user profiles.

Correctness: The actual output mentions enabling notifications and watching entities, similar to the expected output, but lacks specifics about `xwiki.properties`, following pages or users, and the network tab in the user profile.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_023_result.json

Overall_score: 0.7278700368467894

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 0.5

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Correctness: 0.8005535544140704

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the question about enabling notifications for one's own actions, but includes irrelevant links and sources that do not directly contribute to solving the issue.

Faithfulness: The score is 0.50 because the actual output incorrectly claims that switching off the 'Own event filter' allows receiving notifications for your own actions, and mentions a screenshot in the documentation that is not present in the retrieval context.

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which states, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' is ranked first. This directly addresses the input question about enabling notifications for one's own actions, while the irrelevant nodes, discussing RSS feeds, general settings, and technical aspects, are ranked lower.

ContextualRecall: The score is 0.50 because the first sentence is supported by the 1st node in the retrieval context, but the second sentence lacks support as there is no mention of 'Options de filtrage avancées' or similar settings in the user's profile.

CustomContextualRelevancy: The retrieval context confirms that by default, users do not receive notifications for their own actions and mentions the ability to change this in settings, aligning closely with the expected output. However, it does not explicitly mention the 'Options de filtrage avancées' section.

Correctness: The actual output accurately reflects the default notification settings and the ability to disable the filter, matching the expected output. However, it contains extra procedural details and a screenshot not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_024_result.json

Overall_score: 0.7730237483789392

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Correctness: 0.1934056481683713

Reasons:

AnswerRelevancy: The score is 0.89 because the response mostly addresses how to disable notifications, but includes irrelevant links that do not directly help with managing notifications.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant node. The first three nodes are relevant as they discuss 'Choosing the default auto-watched mode', 'Notifications Auto Watch Preferences', and 'Filters' and 'Auto Watch', all of which directly relate to modifying notification settings. The irrelevant node, ranked fourth, discusses 'RSS feeds and notification types', which do not directly relate to disabling automatic notifications.

ContextualRecall: The score is 0.75 because while the 3rd node in the retrieval context supports the sentences about automatic page watching and modifying notification settings, none of the nodes explicitly mention being notified about changes to your work.

CustomContextualRelevancy: The retrieval context mentions auto-watched pages and custom filters preferences, aligning with the expected answer about auto-watching and custom filters. However, it lacks explicit mention of modifying notification settings in the user profile.

Correctness: The actual output discusses notification settings management, but lacks details on 'Suivi automatique des pages' and 'Filtres personnalisés', which are key in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_025_result.json

Overall_score: 0.3390220640949486

Individual_scores:

AnswerRelevancy: 0.45454545454545453

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.17958693002423715

Reasons:

AnswerRelevancy: The score is 0.45 because the response includes several irrelevant source links and introductory phrases that do not directly address the question about configuring automatic user activation. However, it likely contains some relevant information, which is why the score is not lower.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based authentication and bypassing OpenID Connect,' which does not address enabling user accounts automatically upon first login. The second node mentions 'container authentication and OpenID Connect,' but lacks details on automatic user activation. The third node describes 'OpenID Connect authentication and links to configuration guides,' but does not address automatic user activation. The fourth node lists 'dependencies for the OpenID Connect extension,' without information on user account activation. The fifth node focuses on 'LDAP and Active Directory integration,' which is unrelated to the input query.

ContextualRecall: The score is 0.40 because while the 3rd node in the retrieval context mentions 'xwiki.properties' and OpenID Connect, it lacks details on adding or updating properties, saving the file, restarting XWiki, and automatically activating new user accounts on first login.

CustomContextualRelevancy: The retrieval context does not mention adjusting the xwiki.properties file or the specific property oidc.enableUser for automatic user activation via OpenID Connect.

Correctness: The actual output states that automatic activation of user accounts on first login is not supported by OpenID Connect in XWiki, contradicting the expected output, which provides instructions for enabling this feature.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.43045614344122146

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Correctness: 0.3827368606473285

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and directly addresses the question about customizing user profile fields in XWiki using OpenID Connect, with no irrelevant statements.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect provider endpoints' but lacks information on 'user profile field mapping or customization'. The second node talks about 'container authentication and Active Directory' without addressing 'mapping user profile fields from OpenID Connect to XWiki'. The third node describes the 'OpenID Connect Authenticator extension' but omits details on 'how to map user profile fields'. The fourth node lists 'dependencies for the OpenID Connect Authenticator extension' without relating to 'user profile field mapping'. Finally, the fifth node provides 'configuration guides for specific systems' but does not mention 'mapping user profile fields in XWiki using OpenID Connect'.

ContextualRecall: The score is 0.20 because while the 1st node in the retrieval context supports the mention of 'xwiki.properties' and OpenID Connect in sentence 1 of the expected output, the retrieval context lacks details on 'oidc.user.mapping', examples of mapping custom properties, and instructions for saving and restarting XWiki as described in sentences 2 to 5.

CustomContextualRelevancy: The retrieval context does not contain information about customizing user profile fields or using 'oidc.user.mapping' in 'xwiki.properties'.

Correctness: The actual output uses 'xwiki.cfg' instead of 'xwiki.properties', and lists different mappings like 'oidc.userinfo.mapping.username' instead of 'oidc.user.mapping'. It lacks examples using '\${oidc.user.subject}' and does not mention restarting XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5177710101864561

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.3

Correctness: 0.30662606111873647

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the process of configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes source links that are not directly relevant to the specific configuration process, which slightly detracts from the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which discusses 'group synchronization and the use of the 'oidc.groups.mapping' property,' is ranked third. This node is directly relevant to the input query. However, it is preceded by two irrelevant nodes: the first node discusses 'authentication management and OpenID Connect provider endpoints' without mentioning group synchronization, and the second node talks about 'container authentication and other authentication methods' without providing information on configuring group synchronization with OpenID Connect. These irrelevant nodes should be ranked lower than the relevant node to improve the

score.

ContextualRecall: The score is 0.33 because while the 3rd node in the retrieval context covers configuring properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', it lacks details on opening the file, providing examples, replacing placeholders, and saving changes, which are essential parts of the expected output.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using 'oidc.groups.mapping', but lacks details on configuring 'oidc.groups.claim' and 'oidc.groups.separator'. It also doesn't mention editing 'xwiki.properties' or restarting XWiki.

Correctness: The actual output discusses 'oidc.userinfoclaims' and creating a listener for custom synchronization which is not mentioned in the expected output. It lacks instructions on setting 'oidc.groups.claim' and 'oidc.groups.separator' as outlined in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.33988256158381863

Individual_scores:

AnswerRelevancy: 0.6875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.351795369502912

Reasons:

AnswerRelevancy: The score is 0.69 because the response partially addresses the question about limiting file types and sizes for attachments, but the repeated irrelevant source links reduce its overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'language preferences and translations in XWiki', which is unrelated to file upload restrictions. The second node talks about 'various XWiki applications and features, such as administration and PDF export', but does not mention file upload restrictions. The third node is about 'CSS and XSL customization for PDF exports', which does not relate to setting file upload restrictions. The fourth node covers 'various XWiki macros and features like PDF viewer and office document viewer', but does not address file upload restrictions. The fifth node consists of 'image links and references', which are not relevant to configuring file upload restrictions. Finally, the sixth node contains 'image links and references', which do not provide information on setting file upload restrictions.

ContextualRecall: The score is 0.00 because none of the expected output sentences can be attributed to any nodes in the retrieval context, as the context lacks information on configuring MIME types, setting maximum upload size, or any related steps.

CustomContextualRelevancy: The retrieval context does not contain any information related to configuring MIME types or setting maximum attachment size as described in the expected output.

Correctness: The actual output suggests using extensions 'File Size Limit' and 'Allowed Attachments' instead of editing 'xwiki.properties' and XWikiPreferences as in the expected output. It does not mention the MIME type configuration or the exact process for setting upload size in the preferences.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.6119279103919998

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.13823412901866522

Reasons:

AnswerRelevancy: The score is 0.75 because the response partially addresses the question about refusing script rights to a space administrator, but includes several irrelevant links that do not directly answer the specific query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.58 because relevant nodes discussing programming and script rights in XWiki, such as the second node stating 'programming rights can only be granted from the wiki preferences page,' and the third node noting 'ADMIN' is 'not deniable,' are ranked after an irrelevant first node about language preferences and translations. Additionally, irrelevant nodes like the fourth node on checking access rights programmatically and the fifth node about executing code with specific user rights are ranked too closely to relevant nodes, which affects the precision score.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the retrieval context, specifically supported by the 3rd node in retrieval context. Great job!

CustomContextualRelevancy: The retrieval context mentions that programming and admin rights are not deniable, but it does not explicitly state that admin rights cannot be refused, as mentioned in the expected output.

Correctness: Actual output claims script rights can be refused to space admin, contradicting expected output that admin rights can't be refused.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.3538640090330131

Individual_scores:

AnswerRelevancy: 0.7333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.3898507208647453

Reasons:

AnswerRelevancy: The score is 0.73 because the answer partially addresses the question about configuring spaces in XWiki, but the presence of multiple irrelevant source link statements prevents it from being higher.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node is about 'the structure and metadata of a page in XWiki,' which does not address the question of configuring access rights. The second node discusses 'modifying resources using XML and HTTP methods,' which is not relevant to setting access permissions in XWiki. The third node

covers 'managing comments and attachments,' unrelated to configuring team-specific access. The fourth node explains 'the concept of spaces and pages in XWiki,' but does not provide information on setting access rights for teams. The fifth node discusses 'migration actions and code adjustments,' which do not relate to configuring access rights for teams. Finally, the sixth node is about 'configuration display settings in XWiki administration,' which does not help in setting team-specific access rights.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context. The retrieval context lacks information on creating groups, setting rights at the wiki or space level, inheriting rights, and adjusting rights for teams or spaces.

CustomContextualRelevancy: The retrieval context lacks any information related to creating groups, setting rights at wiki or space level, or inheriting rights, which are all key components of the expected output.

Correctness: The actual output covers the visibility and modification restrictions for teams, but lacks details on creating groups and inheriting rights for child pages as specified in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7158095865492499

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Correctness: 0.5662860907240704

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about authenticating users with access tokens, but it includes multiple irrelevant source links that do not contribute to answering the question directly.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because the second node in the retrieval context, which 'explains how to configure authorized applications', is relevant and ranked higher than some irrelevant nodes. However, the first node discusses 'LDAP configuration, which is unrelated to token-based authentication using JWTs', and should be ranked lower. Additionally, the third node 'describes various claims and scopes for access tokens, but does not directly address the process of authenticating users with JWTs', which should also be ranked lower than the relevant fourth node that 'provides a description of the extension that allows authentication using JWTs'.

ContextualRecall: The score is 0.71 because while nodes in the retrieval context provide information about authenticating users with JWTs, enabling the token authenticator, configuring authorized applications, and limitations on user creation, they lack details on generating a JWT token with specific signing key requirements and including the token in the 'Authorization' header.

CustomContextualRelevancy: The retrieval context includes details about configuring authorized applications and JWT usage, matching the expected answer's description of token-based authentication. However, it lacks specific details about JWT claims and the requirement for an Ed25519 key.

Correctness: The actual output correctly mentions using JWT for authentication, setting `xwiki.authentication.authclass`, and creating/updating users based on JWT. However, it introduces OpenID Connect, which is not part of the expected output, and lacks details about JWT claims,

authorized applications, and token usage in headers.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.3253854636303061

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2380270674961223

Reasons:

AnswerRelevancy: The score is 0.71 because the main content addresses the question about authenticating users with access tokens, but it includes irrelevant source links that do not directly relate to the topic, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'model validation strategies in XWiki', which is unrelated to 'token-based authentication or JWT'. The second node talks about 'attachment validation and configuration', which does not pertain to 'user authentication using access tokens'. The third node provides information about 'the admin UI for model validation', which is not relevant to 'authenticating users with access tokens'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, as the context lacks any mention of the key concepts and steps described in the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are central to the expected output.

Correctness: Actual output mentions authenticating users with access tokens but lacks details on JWT, specific claims, Ed25519 signing, configuration steps, and limitations outlined in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.3078174748855923

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.13261913502783948

Reasons:

AnswerRelevancy: The score is 0.71 because the response mostly addresses the question about authenticating users with access tokens, but it includes irrelevant source links that do not directly contribute to answering the query.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'upgrading and downgrading XWiki', which is unrelated to 'user authentication with access tokens'. The second node is about 'features related to users and rights management' without mentioning 'authentication with access tokens'. The third node covers 'rights management and user profiles' but does not include 'authentication using access tokens'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be linked to any nodes in the retrieval context, indicating a complete lack of relevant information in the context to support the expected content.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication or JWT, which are key elements in the expected output.

Correctness: The actual output mentions authentication with access tokens but refers to a different extension (Administration Application) and lacks details on JWT claims, Ed25519 signing, and specific configuration steps outlined in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_mistral2_large

File: qa_001_result.json

Overall_score: 0.36422064452702657

Individual_scores:

AnswerRelevancy: 0.8709677419354839

Faithfulness: 0.8888888888888888

ContextualPrecision: 0.30952380952380953

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.11594342681397676

Reasons:

AnswerRelevancy: The score is 0.87 because the answer is mostly relevant to the question about the bell icon in XWiki, but it includes repeated irrelevant statements about a source link that do not directly address the issue.

Faithfulness: The score is 0.89 because the actual output incorrectly suggests that manual installation of the Notifications Application is possible when the Extension Manager is not functioning, contradicting the retrieval context which states that offline installation of extensions in XWiki is not supported.

ContextualPrecision: The score is 0.31 because relevant nodes in the retrieval context are ranked lower than irrelevant nodes. For example, the third node, which 'mentions various ways to receive notifications in XWiki and references the Notifications Application,' is relevant but ranked after nodes discussing 'installation instructions and prerequisites' (first node) and 'access issues and error messages related to configurable applications' (second node), which are not relevant to the issue of the bell icon not appearing. Additionally, the seventh node, which 'provides information on configuring notification settings in the administration,' is relevant but ranked after nodes that describe 'the content and behavior of the Alert menu' (fourth node) and list 'dependencies for a specific extension' (fifth node), which do not address the issue.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention the `notifications.enabled` setting or `xwiki.properties` file, which are crucial for enabling notifications as described in the expected output.

Correctness: The actual output does not mention enabling notifications in xwiki.properties, which is the key instruction in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6978716227798297

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.875

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Correctness: 0.3122297366789786

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question without any irrelevant information. Great job!

Faithfulness: The score is 0.88 because the actual output incorrectly claims that new users do not receive notifications until they start watching something and enable specific application types, whereas the context clarifies that the whole wiki is watched by default for new users in versions before 15.5, with certain application types like Mention enabled.

ContextualPrecision: The score is 1.00 because the relevant node, which explains 'how to follow a user to receive notifications about their activities', is ranked first. This ensures that the most pertinent information is prioritized, providing an excellent retrieval result!

ContextualRecall: The score is 0.50 because while some sentences, like those about using the 'alert' menu (supported by node 6 in the retrieval context) and following a user (supported by node 1 in the retrieval context), are well-supported, critical details about enabling notifications in 'xwiki.properties' and subscribing to pages are missing in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions following users and watching pages, which aligns with the expected output. However, it lacks details on enabling notifications via xwiki.properties and using the alert menu.

Correctness: The actual output provides reasons for not receiving notifications but misses key details like enabling notifications in xwiki.properties and subscribing to specific pages or users as mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9514770440446026

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.9588622642676154

Reasons:

AnswerRelevancy: The score is 0.75 because the response partially addresses how to enable notifications for one's own actions in XWiki, but includes irrelevant source links that do not directly contribute to solving the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant nodes in the retrieval context, which explain how to enable notifications for one's own actions by mentioning options like 'Advanced filtering options' and 'Own event filter', are ranked higher than the irrelevant nodes. The irrelevant nodes discuss topics such as following a user, administrator settings, and RSS feeds, which do not pertain to the input query. Great job on getting it just right!

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context fully includes the expected answer's information about not receiving notifications for one's own activity by default and the ability to disable this filter in the 'Advanced filtering options' section of the notification settings in the user profile.

Correctness: The actual output aligns with the expected output by correctly stating the default notification setting and the ability to disable the filter in the 'Advanced filtering options' of the notification settings in the user profile.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.4803671637507298

Individual_scores:

AnswerRelevancy: 0.9583333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Correctness: 0.22386964917104518

Reasons:

AnswerRelevancy: The score is 0.96 because the response effectively addresses how to disable notifications in XWiki, but includes a source link that doesn't directly answer the question, slightly affecting its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'following a user and default watch settings', which does not help in disabling notifications for pages you are not interested in. The second node talks about 'selecting the level of details in email notifications and macros for settings', which is not relevant to the problem of disabling notifications. The third node covers 'enabling and disabling the extension and email features', but does not address the specific need to disable notifications for certain pages. All nodes fail to address the input query, resulting in a score of 0.00.

ContextualRecall: The score is 0.50 because only half of the expected output is supported by the nodes in the retrieval context. Specifically, sentences 3 and 4 are supported by nodes 4 and 5 in the retrieval context, while sentences 1 and 2 lack direct support.

CustomContextualRelevancy: The retrieval context mentions default watch settings and user profile settings but lacks details on automatic page watching and deleting custom filters.

Correctness: The actual output focuses on managing notifications broadly and does not mention automatic page watching or custom filters. It introduces the xwiki.properties file and macros, which are

not in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.9003313056717284

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Correctness: 0.8686545006970372

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to configuring OpenID Connect in XWiki, but it slightly deviates by mentioning LDAP Authenticator, which is unrelated to the user's query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node, which contains the property '# oidc.enableUser=true', is ranked first. This directly addresses the need to enable user accounts on first login, ensuring that all irrelevant nodes, such as those discussing 'domain-based instance authentication and session cookie configuration' (second node), 'resetting passwords and different authentication mechanisms' (third node), and others, are ranked lower.

ContextualRecall: The score is 0.60 because while the nodes in the retrieval context support the instructions to enable user accounts by setting the 'oidc.enableUser' property, they do not mention the need to save the file and restart XWiki, nor do they explain the outcome of enabling new user accounts.

CustomContextualRelevancy: The retrieval context contains the exact information found in the expected answer, specifically the property '# oidc.enableUser=true' in the 'xwiki.properties' file, which needs to be uncommented and set to 'true' to enable user accounts on first login via OpenID Connect.

Correctness: The actual output correctly describes setting the oidc.enableUser property to true in the xwiki.properties file and restarting XWiki. However, it lacks guidance on uncommenting the line and uses a different format for adding the property. It does not mention the comments about enabling user accounts, which are present in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.8863405953627065

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.625

ContextualRecall: 0.8

CustomContextualRelevancy: 1.0

Correctness: 0.8930435721762396

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.62 because the first node in the retrieval context is relevant and directly addresses customizing user profile fields by mentioning 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user'. However, several irrelevant nodes are ranked higher than the second relevant node. For example, the second node discusses 'access token claims like profile, email, and address', which are unrelated to the task, and the third node focuses on 'group synchronization and custom claims for id tokens', which do not pertain to mapping user profile fields. These irrelevant nodes should be ranked lower than the eighth node, which is relevant as it lists 'available variables like oidc.user.subject and oidc.user.mail', crucial for mapping additional user profile fields.

ContextualRecall: The score is 0.80 because most of the expected output is supported by the nodes in the retrieval context, particularly regarding the customization and mapping of user profile fields. However, the instruction to save and restart XWiki is not found in the retrieval context, which affects the score.

CustomContextualRelevancy: The retrieval context includes detailed instructions on customizing user profile fields using the `oidc.user.mapping` property in the `xwiki.properties` file, matching the expected output.

Correctness: The actual output accurately describes customizing user profile fields using `oidc.user.mapping` in `xwiki.properties`, includes examples and the need to save configurations, similar to the expected output. However, it omits the specific instruction to restart XWiki to apply changes.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.7450739750657117

Individual_scores:

AnswerRelevancy: 0.9210526315789473

Faithfulness: 1.0

ContextualPrecision: 0.8333333333333334

ContextualRecall: 0.46153846153846156

CustomContextualRelevancy: 0.7

Correctness: 0.5545194239435276

Reasons:

AnswerRelevancy: The score is 0.92 because the response is mostly relevant and provides useful information on configuring group synchronization. However, the repeated and unnecessary 'For example:' statements slightly detract from the overall clarity and focus, preventing a perfect score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.83 because the first two nodes in the retrieval context are relevant, discussing 'oidc.groups.mapping' and 'oidc.groups.separator', which are crucial for configuring group synchronization. However, the third node is less relevant as it focuses on a general description and dependencies of the OpenID Connect extension. The fourth node, discussing authentication management and endpoints, and the fifth node, listing dependencies, are also not directly related to the configuration process. The sixth node is relevant, discussing 'oidc.groups.claim', which is important for the configuration process.

ContextualRecall: The score is 0.46 because while several specific configurations like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` are supported by nodes in the retrieval context, many general instructions such as locating and editing the `xwiki.properties` file, and saving changes are not directly found in the context.

CustomContextualRelevancy: The retrieval context includes details on configuring group synchronization, such as `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, but lacks explicit instructions on locating and editing the `xwiki.properties` file and restarting the XWiki instance.

Correctness: The actual output includes most steps such as configuring the 'oidc.groups.claim' and 'oidc.groups.separator' properties, but introduces additional concepts like 'oidc.groups.prefix' and 'oidc.groups.allowed', which are not in the expected output. Also, it mentions adding the claim 'xwiki_groups' to 'oidc.userinfoclaims' which is not specified in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6261110027809882

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.6666666666666666

CustomContextualRelevancy: 0.3

Correctness: 0.6566660166859294

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and mostly addresses the question about restricting attachments to images or PDFs of a certain size. However, it includes URLs that are not directly related to configuring these specific attachment restrictions, which slightly affects the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.20 because the relevant node, which 'mentions setting limits on the maximum size of an attachment and provides steps to change it', is ranked fifth. This means it is ranked lower than several irrelevant nodes, such as the first node that 'discusses linking to external files and images, which is unrelated to configuring upload restrictions or size limits'.

ContextualRecall: The score is 0.67 because while the retrieval context (nodes) provides detailed steps for configuring the maximum attachment size (as seen in the 5th node), it lacks information on configuring allowed mimetypes and ensuring only images or PDF files can be uploaded, which are crucial parts of the expected output (as mentioned in sentences 1 and 2).

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific details on configuring xwiki.properties and setting size to 10MB.

Correctness: The actual output accurately describes the maximum size configuration, but uses a different MIME type configuration approach by listing specific subtypes rather than a wildcard and includes additional properties not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.6534252060280247
Individual_scores:
AnswerRelevancy: 0.8181818181818182
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 1.0
CustomContextualRelevancy: 1.0
Correctness: 0.10236941798632948
Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the question about denying script rights to a space administrator, but it includes multiple irrelevant source links that do not contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'programming rights and script execution but does not address denying script rights to a space administrator.' Similarly, the second node 'lists various rights and their default states, but it does not mention the inability to deny rights implied by admin rights.' The third node provides 'a tabular view of rights but does not mention the inability to deny script rights to a space administrator.' These irrelevant nodes are ranked higher than any relevant nodes, resulting in a score of 0.00.

ContextualRecall: The score is 1.00 because every part of the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context confirms that rights implied by admin cannot be denied, aligning with the expected output.

Correctness: The actual output suggests denying script rights at the space level, contradicting the expected output that implies rights implied by admin rights cannot be denied in XWiki.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_010_result.json

Overall_score: 0.5975347035283416
Individual_scores:
AnswerRelevancy: 0.8571428571428571
Faithfulness: 1.0
ContextualPrecision: 0.20833333333333331
ContextualRecall: 0.5
CustomContextualRelevancy: 0.3
Correctness: 0.7197320306938592
Reasons:

AnswerRelevancy: The score is 0.86 because the answer mostly addresses the configuration question in XWiki, but it includes multiple source links that do not directly contribute to solving the problem, which slightly detracts from its relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.21 because the relevant nodes in the retrieval context, which provide information on XWiki's permission system and basic rules for setting rights, are ranked lower (sixth and eighth) than several irrelevant nodes. These irrelevant nodes, such as the first node discussing the organization of content and the second node focusing on parent/child relationships, do not address access control or rights management and should be ranked lower.

ContextualRecall: The score is 0.50 because while the retrieval context (node 6) supports the general idea of setting wiki-wide and page-level rights (related to sentences 2, 3, and 4), it lacks specific details about creating groups for each team (sentence 1) and ensuring all teams can view all spaces and only edit their own space (concluding sentences).

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and inheritance of access rights, which partially aligns with setting wiki-wide and space-level rights. However, it lacks specific details on creating groups for each team and setting 'Edit' rights for team-specific spaces as outlined in the expected output.

Correctness: The actual output aligns with the expected output in terms of global view rights and specific edit rights for each team. However, it lacks creating groups for each team and setting wiki-wide rights for XWikiAllGroup. It includes denying edit rights for other teams, which is not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.4249094297882248

Individual_scores:

AnswerRelevancy: 0.8636363636363636

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.18582021509298513

Reasons:

AnswerRelevancy: The score is 0.86 because the response largely addresses the issue of the notification bell not being available, but it includes multiple irrelevant source links that do not contribute to resolving the specific problem mentioned.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'custom notification filters' but does not address enabling notifications in the wiki settings. The second node lists 'various ways to receive notifications' but does not provide information on enabling notifications in the wiki settings. The third node describes 'different notification preferences macros' but does not mention enabling notifications in the wiki settings. The fourth node provides 'links to various notification-related settings' but does not discuss enabling notifications in the wiki settings. The fifth node describes 'the notifications menu and its features' but does not mention enabling notifications in the wiki settings. The sixth node provides 'links to various notification settings' but does not discuss enabling notifications in the wiki settings. The seventh node discusses 'administrator settings for notifications' but does not specifically mention enabling notifications in the wiki settings. The eighth node discusses 'email notifications and their configuration' but does not address enabling notifications in the wiki settings.

ContextualRecall: The score is 0.50 because only the instruction to edit `xwiki.properties` is supported by the 6th node in the retrieval context, while the specific mention of enabling notifications is not found in any node.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via `xwiki.properties` or setting `notifications.enabled` to `true`, which is the key information in the expected output.

Correctness: Actual output lacks the specific configuration detail of setting `notifications.enabled` to `true` in `xwiki.properties` as mentioned in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_012_result.json

Overall_score: 0.31894214189332853
Individual_scores:
AnswerRelevancy: 0.8947368421052632
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.018916009254707832
Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant to the issue of not receiving notifications, but it includes source links that do not directly address the problem, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'image links and references' which are 'not related to notifications or their settings.' Similarly, the second node also focuses on 'image links and references' that 'do not provide any information about enabling notifications.' The third and fourth nodes continue with 'image links and references,' which are 'unrelated to the topic of notifications.' Finally, the fifth node talks about 'permissions and a security module,' which is 'not relevant to enabling or managing notifications.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to enabling notifications or following pages/users as described in the expected output.

Correctness: The actual output does not address enabling notifications in `xwiki.properties` or following pages/users as outlined in the expected output. It instead provides generic troubleshooting steps unrelated to the specific instructions given in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_013_result.json

Overall_score: 0.34662603692377547
Individual_scores:
AnswerRelevancy: 0.8235294117647058
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.2562268097779471
Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the question about enabling notifications for one's own actions, but the inclusion of a source link, which is not directly relevant, slightly detracts from the overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customization of page templates, internationalization, and accessibility', which are unrelated to notification settings. The second node provides information about 'administrating wiki features', but does not mention notification settings or user activity notifications. The third node includes information about 'importing office documents and other extensions', but does not address notification settings for user activities. The fourth node focuses on 'themes, panels, and presentation settings', which are unrelated to notification settings for user activities. The fifth node discusses 'access rights and permissions', which do not relate to enabling notifications for one's own activities.

ContextualRecall: The score is 0.00 because none of the expected output sentences are supported by the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters mentioned in the expected output.

Correctness: Actual output provides general steps for activating notifications, but lacks specific details from expected output about disabling filters in advanced options. No contradiction but insufficient detail match.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.3017603115128699

Individual_scores:

AnswerRelevancy: 0.6111111111111112

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.1994507579661084

Reasons:

AnswerRelevancy: The score is 0.61 because the output contains multiple irrelevant links that do not address the question about disabling notifications, which lowers the score. However, some relevant information might still be present, keeping the score above average.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node is about 'attaching files' and does not mention notifications or settings related to them. The second node discusses 'moving attachments' and does not relate to notifications or their settings. The third node is about 'displaying content and macros', which is unrelated to notification settings. The fourth node is about the 'Distribution Wizard' and does not mention notifications or how to manage them. The fifth node addresses 'linking to various elements' and does not address notifications or their settings. The sixth node discusses 'administration features' and does not relate to notifications or their management. The seventh node is about 'security authorization' and does not mention notifications or how to disable them. The eighth node is about 'importing office documents' and does not relate to notifications or their settings.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about automatic page following or notification settings as described in the expected output.

Correctness: The actual output does not match the expected output. It focuses on managing notification settings in XWiki, whereas the expected output describes automatic page following after significant changes. There are no direct mentions of altering 'Automatisches Verfolgen von Seiten' settings or deleting custom filters as specified in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.402489421396443

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.2149365283786579

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring XWiki for automatic user account activation with OpenID Connect. Great job!

Faithfulness: The score is 0.80 because the actual output incorrectly states that adding the attribute `sessionCookieDomain` to `context.xml` in Tomcat ensures the cookie is set on a top-level domain for all wikis, contradicting the retrieval context which notes this will not work if the wikis have very different domains.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'domain-based authentication and session cookie configuration,' which is not related to enabling automatic user activation with OpenID Connect. The second node focuses on 'container authentication' and mentions OpenID Connect but lacks information about automatic user activation. Subsequent nodes similarly fail to address the specific configuration needed for automatic user activation, such as the third node, which describes the 'OpenID Connect Authenticator extension and its capabilities' without mentioning the required settings.

ContextualRecall: The score is 0.40 because while the 5th node in the retrieval context mentions `xwiki.properties` and `oidc.enableUser`, it does not provide information about updating or adding properties, restarting the XWiki instance, or the automatic activation of new user accounts upon first login via OpenID Connect.

CustomContextualRelevancy: The retrieval context lacks information about modifying `xwiki.properties` for automatic user account activation with OpenID Connect, as specified in the expected output.

Correctness: The actual output discusses user account creation and settings in `xwiki.cfg` and servlet configurations, which are not present in the expected output. It also does not mention updating `xwiki.properties` or the specific property `oidc.enableUser=true`.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.5752349857254923

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.6

CustomContextualRelevancy: 0.2

Correctness: 0.48474324768628685

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the input question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node in the retrieval context, which is the sixth node, is ranked lower than several irrelevant nodes. The first node discusses 'various authentication methods like Container Authentication and Active Directory', which is unrelated to the customization of OpenID Connect user profile fields. The second node focuses on 'domain-based authentication and session cookies', which does not pertain to mapping user profile fields from an OpenID Connect provider. The third node, while mentioning OpenID Connect, primarily discusses 'authentication and synchronization', not the customization of user profile fields. The fourth node provides 'general information about OpenID Connect authentication' without addressing user profile field customization. The fifth node lists 'dependencies for the OpenID Connect extension', which is unrelated to the customization of user profile fields. These irrelevant nodes should be ranked lower than the sixth node, which mentions 'the xwiki.properties file and discusses OpenID Connect endpoints', directly relevant to the input query.

ContextualRecall: The score is 0.60 because while the general use of the xwiki.properties file and the oidc.user.mapping property are supported by the 7th node in the retrieval context, the specific example and detailed instructions are not directly found in the nodes.

CustomContextualRelevancy: The retrieval context mentions OpenID Connect and xwiki.properties but lacks details on customizing user profile fields or using oidc.user.mapping as described in the expected output.

Correctness: The actual output captures the intent of customizing user profile fields using OpenID Connect but lacks specific example mappings provided in the expected output. It includes additional information about installing and configuring the authenticator, which is not in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.5109136541485398

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.20833333333333331

ContextualRecall: 0.3

CustomContextualRelevancy: 0.2

Correctness: 0.35714859155790546

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about setting up group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.21 because relevant nodes in the retrieval context are ranked lower than several irrelevant nodes. The first node discusses 'general OpenID Connect configuration

and endpoints' without addressing group synchronization, and the second node provides 'information about different authentication methods and extensions' without mentioning group synchronization. These should be ranked lower than the sixth node, which mentions 'Group synchronization' and provides details about using claims and mapping properties in the xwiki.properties file. Additionally, the eighth node also discusses 'Group synchronization' and should be ranked higher than the irrelevant nodes.

ContextualRecall: The score is 0.30 because while the retrieval context nodes mention properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, they do not cover the broader steps required for setting up group synchronization between XWiki and an OpenID Connect provider, such as locating and editing the `xwiki.properties` file, or saving changes and restarting XWiki.

CustomContextualRelevancy: The retrieval context mentions group synchronization and the oidc.groups.mapping property, but lacks details on configuring xwiki.properties, setting oidc.groups.claim, oidc.groups.separator, and restarting XWiki.

Correctness: Key facts such as editing `xwiki.properties` and restarting XWiki are present. However, discrepancies include the use of `oidc.userinfoclaims` instead of `oidc.groups.claim`, and lack of details on `oidc.groups.separator`. Additional information like implementing a listener is also present, which contradicts the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6821443407979375

Individual_scores:

AnswerRelevancy: 0.8214285714285714

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.9

CustomContextualRelevancy: 0.5

Correctness: 0.7047708066923872

Reasons:

AnswerRelevancy: The score is 0.82 because the main content of the output is relevant and addresses the question about configuring upload restrictions, but the repeated source links do not directly contribute to answering the question, slightly lowering the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node in the retrieval context, ranked sixth, 'explains how to set the maximum size of an attachment and mentions MIME type restrictions, directly relevant to the input question.' However, it is ranked lower than five irrelevant nodes that discuss topics like 'trusted domains and PDF export customization,' 'ODT and PDF export features,' 'various configuration options like WebDAV and PDF export,' 'PDF and RTF export issues,' and 'links to various attachment-related topics but does not directly address MIME type or size restrictions for uploads.' These irrelevant nodes should be ranked lower than the relevant node.

ContextualRecall: The score is 0.90 because most sentences in the expected output are supported by the 5th node in the retrieval context, which discusses attachment size limits, MIME type restrictions, and provides steps for changing the maximum upload size. However, the specific configuration of MIME types is not explicitly mentioned in the nodes, leading to a slightly lower score.

CustomContextualRelevancy: The retrieval context partially covers the expected answer by mentioning size limits and MIME type restrictions, but lacks specific details on configuring MIME types in 'xwiki.properties' and setting the maximum upload size to 10 MB.

Correctness: Actual output includes all factual information but uses 'Attachment Validation Application' instead of editing 'xwiki.properties' directly for MIME types, which is additional and slightly changes the approach, though context and meaning are maintained.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.31632416752050735

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.09794500512304401

Reasons:

AnswerRelevancy: The score is 0.80 because the answer mostly addresses the question about denying script rights to a domain administrator, but it includes multiple irrelevant source links that do not directly answer the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'form validation', which does not relate to 'rights management or denying script rights in XWiki'. The second node talks about 'template providers and page creation', which is unrelated to the topic. The third node is about 'configuration classes and sheets', which is not relevant to 'denying script rights'. The fourth node focuses on 'field validation in registration forms', which does not pertain to the subject. The fifth node, although discussing 'access rights and context setup', does not address 'denying script rights to a space administrator'.

ContextualRecall: The score is 0.00 because the expected output sentence about XWiki's admin rights cannot be linked to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention anything about rights in XWiki or their implications, which is the core information in the expected output.

Correctness: The actual output provides steps to deny script rights, contradicting the expected output which states this is not supported in XWiki.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.42034984821806365

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.7720990893083821

Reasons:

AnswerRelevancy: The score is 0.75 because the output provides relevant information about configuring team access in XWiki, but it includes several irrelevant source links that do not directly address the specific configuration question, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses XWiki's database efficiency and programming features, which are not relevant to configuring access controls for teams. The second node is about the Notifications Application in XWiki, which does not relate to setting access rights for teams. The third node describes an application for administrating wiki features but does not provide specific guidance on configuring access controls for teams. These irrelevant nodes should be ranked lower than any relevant nodes.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks information about creating groups, setting rights, and inheriting rights, which are key elements in the expected output.

Correctness: The actual output covers key facts like hierarchical structure and team-specific editing rights. However, it misses 'wikiweiten Rechten' and 'Vererben von Rechten' from expected output, and introduces 'Überprüfung und Anpassung' which is not in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.8255386179923728

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.7941407988633283

Reasons:

AnswerRelevancy: The score is 0.91 because the response is mostly relevant and addresses the issue of the notification bell not being available, but it includes source links that do not directly contribute to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 0.25 because the relevant node, which mentions the 'notifications.enabled' property in 'xwiki.properties' and is directly relevant to enabling notifications, is ranked fourth. However, it is preceded by three irrelevant nodes. The first node discusses 'custom notification filters and their behavior', the second node provides 'links and descriptions of various notification preferences', and the third node covers 'technical details about notification filters and hidden pages', none of which are directly related to enabling notifications in the wiki. This misordering of nodes results in a lower score.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context explicitly mentions enabling notifications by setting 'notifications.enabled' in 'xwiki.properties' to 'true', which matches the expected output.

Correctness: The actual output accurately includes the key point about setting 'notifications.enabled' to 'true' in 'xwiki.properties', aligning with the expected output. However, it contains additional troubleshooting steps and sources not present in the expected output, which do not affect factual

correctness but introduce extra information.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.6560047699568734

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0.7222222222222222

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Correctness: 0.27491750863012926

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant to the issue of not receiving notifications, but it includes links that are not directly related to solving the problem, which slightly detracts from its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.72 because relevant nodes in the retrieval context, such as the first node mentioning 'Enable and disable the extension' and 'Enabling notifications for all users by default', are ranked higher. However, the second node, which focuses on 'Choosing the default frequency of notification emails' and 'Create global notification filters', is not directly related to the issue of not receiving notifications and should be ranked lower. Similarly, the fourth node discussing 'Get a notification RSS feed' and 'Notifications Macro' is not directly related to enabling notifications or following users, which affects the score.

ContextualRecall: The score is 0.75 because while several aspects of the expected output, such as enabling notifications and following users or pages, are supported by nodes in the retrieval context, specific details like setting 'xwiki.properties' to 'true' and the 'network' tab in the user profile are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks specific steps like setting 'notifications.enabled' to 'true' in 'xwiki.properties' and using the alert menu.

Correctness: The actual output mentions checking notification settings and following pages, but lacks details on setting 'notifications.enabled' in 'xwiki.properties', following users, and managing alerts via the bell icon.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8220538715666449

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Correctness: 0.6434343405109801

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant and helpful in addressing how to enable notifications for one's own actions. However, it includes an irrelevant link to XWiki syntax documentation, which slightly detracts from the focus of the answer.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which states, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' is ranked first, perfectly addressing the input question about not receiving notifications for one's own activity and how to change this setting. The irrelevant nodes, which discuss topics like general notification settings, RSS feeds, and unrelated notification preferences, are all ranked lower, ensuring precise and relevant information is prioritized.

ContextualRecall: The score is 0.50 because while the first sentence is supported by the 1st node in the retrieval context, the information about disabling the filter in 'Options de filtrage avancées' is not found in any of the nodes.

CustomContextualRelevancy: The retrieval context confirms that by default, users do not receive notifications for their own actions and mentions the ability to change this setting by switching off the 'Own event filter'. However, it does not explicitly mention 'Options de filtrage avancées' as the section to change this setting, which slightly reduces the completeness.

Correctness: The actual output correctly states that users do not receive notifications for their own actions by default, aligning with the expected output. However, it lacks mentioning that notifications are received if the event explicitly targets the user, and it provides additional information on how to change the settings, which is not in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.48025506237387555

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 0.9090909090909091

ContextualPrecision: 0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Correctness: 0.1724394651523444

Reasons:

AnswerRelevancy: The score is 0.75 because the response partially addresses the question about disabling notifications, but includes several irrelevant statements such as source links and macro pages that do not directly provide the necessary steps or information.

Faithfulness: The score is 0.91 because the actual output incorrectly claims that users do not receive notifications about their own activities by default, whereas the retrieval context clarifies that notifications are not received unless the event explicitly targets them.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'various filters and settings for notifications' but does not address how to disable notifications or mention the 'Suivi automatique des pages' feature. Similarly, the second node lists 'different notification preferences and macros' without mentioning how to disable notifications or the 'Suivi automatique des pages' feature. The third node provides 'links and descriptions of notification settings' but lacks information on disabling notifications or the 'Suivi automatique des pages' feature. The fourth node mentions 'RSS feeds and notification types' but does not provide information on disabling notifications or the 'Suivi automatique des pages' feature. Finally, the fifth node lists 'various

ways to receive notifications, such as RSS feeds and mentions' but does not explain how to disable notifications or the 'Suivi automatique des pages' feature.

ContextualRecall: The score is 0.75 because most of the expected output is supported by nodes in the retrieval context, such as the mention of 'Suivi automatique des pages' and 'Filtres personnalisés' in node 2. However, the retrieval context does not mention being notified of changes on one's work, which affects the score.

CustomContextualRelevancy: The retrieval context mentions auto watch and notification settings but lacks details on modifying 'Suivi automatique des pages' and removing filters in 'Filtres personnalisés'.

Correctness: The actual output lacks key information such as automatic tracking of pages after significant changes and adjusting settings in 'Filtres personnalisés' in user profile notifications settings, which are present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.4073999801965235

Individual_scores:

AnswerRelevancy: 0.9411764705882353

Faithfulness: 0.8

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.1

Correctness: 0.20322341059090548

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant to configuring OpenID Connect in XWiki, with only a minor irrelevant mention of LDAP Authenticator, which slightly detracts from a perfect score.

Faithfulness: The score is 0.80 because the actual output incorrectly claims that restarting the XWiki server is necessary after changes in xwiki.cfg, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based authentication and session cookies', which is unrelated to enabling user accounts automatically with OpenID Connect. The second node provides 'information about OpenID Connect endpoints and headers', but does not mention enabling user accounts automatically. The third node describes 'installation instructions and limitations of OpenID Connect', but does not address the automatic activation of user accounts. The fourth node lists 'dependencies for the OpenID Connect extension', which is not relevant to configuring automatic user activation. The fifth node discusses 'LDAP and Active Directory integration', which is not related to OpenID Connect or automatic user activation. The sixth node provides 'general information about the OpenID Connect Authenticator', but does not mention automatic user activation. The seventh node describes 'container authentication and various authentication methods', but does not address the specific configuration for automatic user activation with OpenID Connect. All these irrelevant nodes are ranked higher, resulting in a score of 0.00.

ContextualRecall: The score is 0.40 because while the retrieval context mentions 'xwiki.properties' and 'oidc.enableUser', which relate to sentences about activating user accounts using OpenID Connect, it lacks information on adding or updating the property, saving the file, restarting XWiki, and explicitly guaranteeing automatic activation of new user accounts.

CustomContextualRelevancy: The retrieval context mentions the xwiki.properties file but lacks specific details about setting oidc.enableUser to true for automatic user activation.

Correctness: The actual output references `xwiki.authentication.createuser` in `xwiki.cfg` instead of `oidc.enableUser` in `xwiki.properties`, missing key expected details and altering context.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_026_result.json

Overall_score: 0.4920638763484731
Individual_scores:
AnswerRelevancy: 0.7777777777777778
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.4
CustomContextualRelevancy: 0.2
Correctness: 0.574605480313061
Reasons:

AnswerRelevancy: The score is 0.78 because the response provides relevant information about customizing user profile fields in XWiki using OpenID Connect, but it includes several source links and an introductory phrase that do not directly address the input question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'container authentication, Active Directory, and LDAP', which does not mention customizing OpenID Connect user profile fields. The second node is about 'domain-based authentication and session cookie configuration', unrelated to mapping OpenID Connect user profile fields. The third node lists 'endpoints and customization options for OpenID Connect', but lacks information on mapping user profile fields in `xwiki.properties`. Subsequent nodes continue to focus on unrelated aspects like general information, dependencies, and installation instructions for the OpenID Connect Authenticator, without addressing the specific query about mapping user profile fields.

ContextualRecall: The score is 0.40 because while the 2nd node in the retrieval context supports the mention of 'xwiki.properties' and restarting XWiki (related to sentences 1 and 5), there is no specific information in the retrieval context about 'oidc.user.mapping', mapping additional fields, or the example mappings provided in sentences 2, 3, and 4.

CustomContextualRelevancy: The retrieval context mentions OpenID Connect and xwiki.properties but lacks specific details on customizing user profile fields or using oidc.user.mapping as described in the expected output.

Correctness: The actual output provides a correct overview of the process but lacks specific mapping examples and the mention of restarting XWiki as described in the expected output. It also introduces additional steps like installing extensions which are not mentioned in the expected output.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5947823586918446
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 0.25
ContextualRecall: 0.4444444444444444
CustomContextualRelevancy: 0.6
Correctness: 0.27424970770662277

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.25 because the relevant node in the retrieval context, which 'directly mentions group synchronization, the use of claims, and the oidc.groups.mapping property', is ranked fourth. It should be ranked higher than the first three nodes, which discuss 'general OpenID Connect provider configuration and endpoints', 'container authentication and other authentication methods', and 'domain-based authentication and session cookies', none of which are related to group synchronization.

ContextualRecall: The score is 0.44 because while the retrieval context aligns with group synchronization and configuration properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', it lacks specific details such as opening the 'xwiki.properties' file with a text editor, replacing placeholders with actual values, and saving and restarting XWiki.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using oidc.groups.mapping, but lacks specific details about configuring oidc.groups.claim and oidc.groups.separator as outlined in the expected output.

Correctness: The actual output mentions adding 'xwiki_groups' to 'oidc.userinfoclaims' and does not address 'oidc.groups.claim' or 'oidc.groups.separator', which are critical in the expected output. It also suggests configurations not covered in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3258229387936233

Individual_scores:

AnswerRelevancy: 0.6842105263157895

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.27072710644595044

Reasons:

AnswerRelevancy: The score is 0.68 because the response partially addresses the question about restricting file uploads to images or PDFs of a certain size, but it includes multiple irrelevant source links that do not directly address the query, which lowers the score.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For instance, the first node discusses 'language preferences and translations in XWiki,' which is unrelated to 'file upload restrictions or MIME types.' Similarly, the second node talks about 'various XWiki applications like tour guides and administration,' which does not mention 'file uploads or size restrictions.' None of the nodes provide relevant information on configuring file upload settings, leading to a low score.

ContextualRecall: The score is 0.00 because none of the expected output sentences can be attributed to any nodes in the retrieval context, as there is no mention of file size restrictions, MIME type configurations, or steps for setting maximum upload size in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to configuring MIME types or attachment size limits as specified in the expected output.

Correctness: Actual output lacks specific steps such as modifying `xwiki.properties` and setting maximum upload size in XWikiPreferences. It only provides general guidance without detailed instructions found in expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.601640050736683

Individual_scores:

AnswerRelevancy: 0.72727272727273

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.1825675771473703

Reasons:

AnswerRelevancy: The score is 0.73 because the main content addresses the question about refusing script rights, but the repeated source links do not contribute to answering the question, which affects the score negatively.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because the relevant nodes in the retrieval context are not consistently ranked higher than the irrelevant nodes. The first node discusses 'language preferences and translations in XWiki', which is unrelated to the input, yet it is ranked higher than the second node that 'discusses programming and script rights', which is relevant. Similarly, the third node about 'interacting with the XWiki RESTful API' is ranked higher than the fourth node that 'explains the script right and its default settings', which is also relevant. This inconsistency in ranking relevant nodes higher than irrelevant ones results in a moderate score.

ContextualRecall: The score is 1.00 because the expected output is perfectly aligned with the information from the retrieval context, specifically the 5th node, which confirms that ADMIN rights are not deniable. Great job maintaining accuracy and relevance!

CustomContextualRelevancy: The retrieval context mentions that administrator rights imply other rights, but it does not explicitly state that administrator rights cannot be refused, which is the key information in the expected output.

Correctness: The actual output provides steps to deny script rights, contradicting the expected output's statement that admin rights cannot be denied. It inaccurately implies that script rights can be refused, which is not supported as per the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.4368608017814643

Individual_scores:

AnswerRelevancy: 0.9324324324324325

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.6887323782563531

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to configuring team space permissions in XWiki, but it includes several source links that are not directly related to the task, slightly lowering the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question about configuring access rights in XWiki. For example, the first node contains 'XML metadata about a page', which is not relevant to configuring access rights. Similarly, the second node discusses 'sending representations and modifying resources via XML', which does not address access control settings. Each subsequent node, like the third one about 'adding, editing, and managing comments', also fails to relate to the topic of access rights configuration in XWiki.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, as the context lacks any mention of steps for achieving access control, creating groups, setting rights, or ensuring visibility and modification permissions in XWiki.

CustomContextualRelevancy: The retrieval context does not contain any information related to creating groups, setting rights at wiki or space levels, or inheriting rights for child pages as described in the expected output.

Correctness: The actual output covers most points but uses different terminology and adds extra details like API usage. Key facts like page inheritance and specific rights settings are not directly mentioned.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7209701682174745

Individual_scores:

AnswerRelevancy: 0.7058823529411765

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.11993865636367071

Reasons:

AnswerRelevancy: The score is 0.71 because the main content of the output addresses the question about authenticating users with access tokens, but the inclusion of multiple irrelevant source links prevents the score from being higher.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because relevant nodes in the retrieval context, such as the second node discussing 'the configuration of authorized applications' and the fourth node detailing 'JWT token requirements and signing process', are ranked lower than irrelevant nodes. For instance, the first node, which is about 'LDAP configuration and does not mention token-based authentication or JWT', is ranked higher than these relevant nodes. This misordering affects the precision score, as irrelevant nodes are not ranked lower than relevant ones.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context contains all the necessary information from the expected answer, including the use of JWT tokens with specific claims, Ed25519 key signing,

configuration steps in 'xwiki.properties', and limitations on user creation and updates.

Correctness: The actual output describes the OpenID Connect Authenticator, which differs from the expected output's Token-based authentication for the LLM Application extension, including different setup details and token requirements.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.2711226236338978

Individual_scores:

AnswerRelevancy: 0.5555555555555556

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.07118018624783105

Reasons:

AnswerRelevancy: The score is 0.56 because the response includes several irrelevant links to XWiki extensions and applications that do not directly address the question about authenticating users with access tokens. However, it still contains some relevant information related to the topic, which is why the score is not lower.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node is about 'Model Validation' and does not mention anything about token-based authentication or JWT. The second node discusses 'Attachment Validation' and configuration, which is unrelated to user authentication with access tokens. The third node is about the 'Admin UI for Model Validation' and does not relate to authentication processes or access tokens. The fourth node is about setting up fields and validation for user registration, not about authenticating users with access tokens. The fifth node describes installation steps for a UI and does not mention authentication or access tokens.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are crucial elements of the expected answer.

Correctness: The actual output does not provide specific information or alignment with the expected output's details on token-based authentication in XWiki.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.2386002809935923

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 0.75

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.18160168596155368

Reasons:

AnswerRelevancy: The score is 0.50 because the response partially addresses the question about authenticating users with access tokens, but includes multiple irrelevant statements such as repeated source links and a general offer for further assistance that do not directly contribute to answering the question.

Faithfulness: The score is 0.75 because the actual output incorrectly claims that XWiki has modules for managing user authentication, while the retrieval context clarifies that the Authentication API in XWiki is not yet implemented.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'upgrading XWiki' and lacks any mention of user authentication or token-based authentication. The second node focuses on 'features of the Administration Application,' which does not relate to token-based user authentication. The third node, although discussing the Security Module, explicitly states that 'The Authentication API is not yet implemented,' which is unrelated to token-based authentication. The fourth node is about 'managing rights' without mentioning token-based authentication or JWT. Finally, the fifth node addresses 'rights management constants and policies,' which are unrelated to authenticating users with access tokens. Therefore, irrelevant nodes are ranked higher than any relevant content, resulting in a score of 0.00.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information about token-based authentication or JWT, which are central to the expected output.

Correctness: The actual output fails to mention the specific JWT-based authentication process and details such as JWT claims, signing with Ed25519, or configuration steps mentioned in the expected output. Instead, it vaguely suggests checking documentation and exploring custom extensions.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_mixtral-8x22b

File: qa_001_result.json

Overall_score: 0.32705001693834407

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.18452232385228687

Reasons:

AnswerRelevancy: The score is 0.78 because the output contains useful information related to XWiki, but it includes multiple irrelevant statements that do not directly address the specific issue of the missing bell icon.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'installation instructions for extensions', which does not help with enabling notifications. The second node describes 'functionality of the Alert menu', not addressing the issue of enabling notifications. The third node lists 'dependencies', offering no information on enabling notifications. Finally, the fourth node talks about 'customizing notification templates', unrelated to the

problem of enabling notifications.

ContextualRecall: The score is 0.00 because the expected output about enabling notifications and setting `notifications.enabled` in `xwiki.properties` is not supported by any information in the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any mention of enabling notifications via the 'notifications.enabled' setting in 'xwiki.properties'.

Correctness: The actual output provides troubleshooting steps for the Alerts and Notifications Applications, but does not mention enabling notifications in `xwiki.properties` as specified in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6779018897754411

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Correctness: 0.3174113386526465

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about not receiving notifications in XWiki. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.75 because the relevant nodes in the retrieval context are mostly ranked higher than the irrelevant ones. However, the second node, which discusses 'filters and technical settings that do not directly address the issue of not receiving notifications or how to enable them,' and the third node, which focuses on 'RSS feeds and notification macros, which are not directly related to enabling notifications or ensuring they are received,' are ranked higher than the fourth node. The fourth node outlines 'different ways to receive notifications in XWiki, including the Notifications Application and mentions,' which is relevant to the input, and should be ranked higher.

ContextualRecall: The score is 0.50 because half of the expected output sentences are supported by nodes in the retrieval context, such as the details about using the 'alert' menu and following users, while the other half, like enabling notifications in `xwiki.properties` and subscribing to pages, are not directly supported by any nodes.

CustomContextualRelevancy: The retrieval context includes information on following users and notification settings but lacks details on enabling notifications in xwiki.properties and watching pages.

Correctness: The actual output correctly advises enabling notifications settings but lacks instructions about subscribing to pages or users. It mentions default settings for new users and customization, which is additional information not found in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9751172472608903

Individual_scores:

AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 1.0
CustomContextualRelevancy: 1.0
Correctness: 0.8507034835653414

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because the relevant nodes in the retrieval context are perfectly ranked higher than the irrelevant nodes. The first and second nodes directly address the issue of not receiving notifications for one's own activity by mentioning options to disable filters in notification settings, while the third, fourth, and fifth nodes discuss unrelated topics such as administrator settings and RSS feeds, which do not pertain to the specific issue.

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context contains all the information from the expected output, including the default behavior of not receiving notifications about one's own activity unless explicitly targeted, and the ability to disable this filter in the 'Advanced filtering options' section of the notification settings.

Correctness: The actual output correctly states that by default, you won't receive notifications for your own actions and describes how to change this setting. It mentions the 'Own Event Filter' and suggests it might be in 'Advanced filtering options,' aligning closely with the expected output. However, it lacks explicit mention of the user profile location for the settings.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5775503012533487
Individual_scores:
AnswerRelevancy: 0.8666666666666667
Faithfulness: 0.875
ContextualPrecision: 0.2
ContextualRecall: 1.0
CustomContextualRelevancy: 0.4
Correctness: 0.12363514085342604

Reasons:

AnswerRelevancy: The score is 0.87 because the response effectively addresses how to disable notifications in XWiki, but includes irrelevant source links that do not directly answer the question.

Faithfulness: The score is 0.88 because the claim implies notifications are received only after changes to settings, whereas the context states notifications are received after enabling notification types.

ContextualPrecision: The score is 0.20 because the relevant node, ranked fifth, correctly addresses changing notification settings by mentioning that 'administrators can change the default auto-watched mode and users can change their notification settings.' However, it is ranked lower than four irrelevant nodes. For example, the first node discusses 'how to get an RSS feed of notifications,' which is unrelated to disabling notifications. Similarly, the second node talks about 'following a user and default watch settings,' which does not provide information on how to disable notifications.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context mentions default watch settings and user profile settings but lacks details about automatic page watching and deleting notification filters.

Correctness: The actual output describes managing notifications in XWiki but lacks the specific details about automatic page watching, changing settings, and deleting filters described in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.9338159333409104

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.8

CustomContextualRelevancy: 0.9

Correctness: 0.9798186769685389

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant and provides useful information about configuring OpenID Connect in XWiki. However, it includes an irrelevant statement about LDAP authentication, which slightly detracts from its focus on the specific OpenID Connect configuration issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions the property 'oidc.enableUser=true', is ranked first. This directly addresses the configuration needed to enable user accounts on first login, ensuring that irrelevant nodes discussing topics like domain-based instances, cookie configurations, and LDAP are ranked lower.

ContextualRecall: The score is 0.80 because the retrieval context nodes cover most of the expected output, including enabling user accounts on first login and the relevant property settings, but they do not mention the need to save the file and restart the XWiki instance.

CustomContextualRelevancy: The retrieval context includes the commented line '#oidc.enableUser=true', which matches the expected answer's requirement to uncomment and set this property to true. However, the context does not explicitly mention saving the file and restarting XWiki.

Correctness: The actual output accurately describes adjusting the xwiki.properties file, including setting oidc.enableUser to true and restarting XWiki, aligning with the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.918798554870337

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.8

CustomContextualRelevancy: 1.0
Correctness: 0.8127913292220222

Reasons:

AnswerRelevancy: The score is 0.90 because the response effectively addresses the main question about customizing user profile fields in XWiki using OpenID Connect, but includes a source link that is not directly relevant to the query.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which states 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user' and provides an example of mapping, is ranked first. This directly addresses the input question about customizing user profile fields. The irrelevant nodes, discussing topics like 'access tokens and scopes', 'configuring OpenID Connect endpoints', 'group synchronization', and 'in-wiki configuration', are all ranked lower, ensuring a perfect contextual precision score. Great job on the ranking!

ContextualRecall: The score is 0.80 because most of the expected output is supported by the 3rd node in the retrieval context, which provides detailed examples and instructions on mapping user profile fields. However, the retrieval context lacks information about saving the 'xwiki.properties' file and restarting XWiki, which affects the completeness of the match.

CustomContextualRelevancy: The retrieval context includes detailed instructions on mapping user profile fields using the `oidc.user.mapping` property in the `xwiki.properties` file, matching the expected output.

Correctness: The actual output closely aligns with the expected output but includes additional details like extra placeholders (`\${oidc.user.preferredUsername}`, `\${oidc.user.email}`) which are not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.847586848122358

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 1.0

Correctness: 0.5855210887341478

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant, addressing the specific question about configuring group synchronization between XWiki and an OpenID Connect provider without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions; the actual output aligns perfectly with the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first three nodes in the retrieval context provide crucial information on configuring group synchronization, such as the `oidc.groups.claim` property and `oidc.groups.mapping`, which are essential for the process. The irrelevant nodes, ranked fourth and fifth, discuss installation instructions and authentication management, which are not directly related to the group synchronization setup. Great job on achieving perfect precision!

ContextualRecall: The score is 0.50 because while the nodes in retrieval context provide specific examples and mention properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, they do not cover the general instructions found in the expected output.

CustomContextualRelevancy: The retrieval context contains all necessary details for configuring group synchronization, including setting the 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' properties, as well as saving changes and restarting XWiki.

Correctness: The actual output includes additional information on OpenID Connect provider support and claim setup, which is not part of the expected output. It correctly mentions configuring `xwiki.properties` for group mapping and separator but uses different claim names and properties like `oidc.userinfoclaims` not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6388471701621393

Individual_scores:

AnswerRelevancy: 0.9473684210526315

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.6

CustomContextualRelevancy: 0.3

Correctness: 0.6523812665868712

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant and provides clear guidance on configuring attachment settings, but it includes an unnecessary link to basic XWiki concepts, which slightly detracts from its focus.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, ranked third, 'explains how to set the maximum size of an attachment in XWikiPreferences, which is directly relevant to configuring the maximum upload size.' However, irrelevant nodes ranked first and second 'discuss linking to external files and images, but do not address configuring upload restrictions or size limits' and 'provide a general overview of XWiki's structure and features, but do not mention attachment upload settings or size restrictions,' respectively, which should be ranked lower than the relevant node.

ContextualRecall: The score is 0.60 because while nodes in the retrieval context support specific steps for configuring mimetypes and setting the maximum upload size (as seen in nodes 3 and 4), the introductory and explanatory sentences, as well as headings, do not have direct support from any nodes.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specifics on configuring 'xwiki.properties' and setting size to 10MB.

Correctness: Actual output uses the Attachment Validation Application instead of editing the xwiki.properties file for mimetype restrictions. Size configuration steps align well with expected output, and no contradictions were noted.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.6296984929444197

Individual_scores:

AnswerRelevancy: 0.6818181818181818

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.09637277584833669

Reasons:

AnswerRelevancy: The score is 0.68 because the answer partially addresses the question of denying script rights to a space administrator, but it includes multiple irrelevant source links and a generic label that do not contribute to the solution.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses programming rights and script execution but does not address the denial of script rights to a space administrator. The second node lists various rights and their properties but does not mention the inability to deny script rights to a space administrator. The third node explains the script right and its default settings but does not state that denying script rights to a space administrator is unsupported. The fourth node contains technical code snippets related to access checks but does not provide information on denying script rights to a space administrator. Finally, the fifth node discusses converting rights and excluding spaces but does not mention the denial of script rights to a space administrator.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the information provided by the node in the retrieval context. Great job!

CustomContextualRelevancy: The expected answer is fully supported by the retrieval context, which states that rights implied by admin cannot be denied.

Correctness: The actual output contradicts the expected output by suggesting that the script right can be denied, whereas the expected output states it cannot be denied due to admin rights implications.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7627337544134255

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.875

CustomContextualRelevancy: 0.5

Correctness: 0.7347358598138862

Reasons:

AnswerRelevancy: The score is 0.80 because the response largely addresses the configuration question in XWiki, but it includes multiple irrelevant source links that do not directly contribute to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output perfectly aligns with the retrieval context. Great job!

ContextualPrecision: The score is 0.67 because the relevant nodes in the retrieval context are mostly ranked higher than the irrelevant nodes. However, the second node, which focuses on the 'parent/child relationship and nested spaces concept', and the third node, which discusses 'string formatting and OpenID Connect properties', are ranked higher than the fourth node, which provides information on

XWiki's permission system, directly relating to the configuration of access. This affects the precision score as these irrelevant nodes should be ranked lower.

ContextualRecall: The score is 0.88 because most of the expected output sentences are supported by the nodes in the retrieval context, such as nodes 4 and 6 aligning with creating groups and setting rights. However, there is a lack of specific mention in the nodes regarding adjusting rights for additional teams or spaces, which slightly affects the score.

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and inheritance of access rights, which partially aligns with the expected output's steps on setting wiki-wide and space-level rights. However, it lacks specific details about creating groups for each team and setting edit rights specifically for team spaces.

Correctness: The actual output correctly details creating groups, setting view and edit rights, and inheriting rights. However, it adds an extra step about administering rights and misses mentioning setting 'View' right specifically for 'XWikiAllGroup' and selecting 'Rights: Page & Children' option.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.4807301982681202

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.21771452294205443

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and addresses the input without any irrelevant statements. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node in the retrieval context, which discusses 'editing the `xwiki.properties` file', is ranked last (6th). This node is relevant to enabling notifications, as suggested in the input. However, it is ranked lower than irrelevant nodes that discuss 'custom notification filters', 'alert menu and notification display', 'various notification methods in XWiki', 'links to notification-related settings', and 'notification settings for administrators', none of which address enabling notifications in the wiki settings.

ContextualRecall: The score is 0.50 because while the 5th node in the retrieval context mentions editing the xwiki.properties file, there is no mention of setting 'notifications.enabled' to 'true', which is a key detail in the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications through the `xwiki.properties` file or setting `notifications.enabled` to `true`.

Correctness: The actual output provides troubleshooting advice and references XWiki sources but lacks specific instructions on enabling notifications in xwiki.properties as stated in the expected output. It also includes irrelevant information about contacting support.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3478060452718996
Individual_scores:
AnswerRelevancy: 0.9090909090909091
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.17774536254048864
Reasons:

AnswerRelevancy: The score is 0.91 because the response effectively addresses the issue of not receiving notifications, but it includes an unrelated source link, which slightly detracts from its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node contains 'a list of image links and references, which are not related to enabling notifications or any settings', and similar irrelevance is found in the second and third nodes. Therefore, relevant nodes are not ranked higher than irrelevant nodes.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context contains image links and references, which do not match any information from the expected output about enabling notifications and following users or pages.

Correctness: The actual output lacks factual alignment with the expected output, omitting critical steps like modifying 'xwiki.properties' and following pages or users. It instead provides generic steps not present in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_013_result.json

Overall_score: 0.35971025193261436
Individual_scores:
AnswerRelevancy: 0.8181818181818182
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.34007969341386807
Reasons:

AnswerRelevancy: The score is 0.82 because the response is mostly relevant to the question about enabling notifications for one's own activity, but it includes irrelevant links that do not directly address the query.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining accuracy and alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customization of page templates, internationalization, and other features unrelated to notification settings', the second node is about 'the administration application for managing wiki features, which does not relate to user notification settings', and the third node covers 'importing office documents, icon themes, and other features, but does not mention notification settings or user activity'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters, which are central to the expected output.

Correctness: The actual output suggests checking notification settings for own actions, which aligns partially with the expected output's mention of advanced filter options. However, it lacks clarity on disabling filters and implies uncertainty about the feature's existence. It also adds unsupported steps for accessing settings and consulting an administrator.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.31327228419792835

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.21296703852090376

Reasons:

AnswerRelevancy: The score is 0.67 because while the output provides some relevant information about disabling notifications, it includes multiple irrelevant links to unrelated XWiki pages, which detracts from its overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files', which does not relate to 'notification settings'. The second node is about 'moving attachments', unrelated to 'managing or disabling notifications'. The third node lists 'display and content macros', not pertaining to 'notification settings'. The fourth node is about 'administrating wiki features', not relevant to 'notification settings'. The fifth node discusses 'permissions and security modules', unrelated to 'notification settings'. Lastly, the sixth node lists 'extensions like Office Importer and Syntax Highlighting', which do not relate to 'notification settings'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to automatic page following or notification settings as mentioned in the expected output.

Correctness: The actual output does not mention automatic following of pages with significant changes, nor the deletion of filters, which are key facts in the expected output. Instead, it describes general notification settings and steps to deactivate them, which is additional information not aligning with the expected context.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.42078766977487914

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.32472601864927464

Reasons:

AnswerRelevancy: The score is 0.80 because the output mostly addresses the query about configuring XWiki for automatic user activation with OpenID Connect. However, some parts of the output include comments and configuration settings that do not directly answer the question, which prevents a higher score.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'domain-based authentication and session cookie configuration', which does not relate to 'enabling user accounts upon first login using OpenID Connect'. The second node provides a 'general description of OpenID Connect authentication' without addressing 'the specific configuration needed to automatically activate user accounts'. The third node offers 'metadata about the OpenID Connect Authenticator extension', which is unrelated to 'user account activation'. Finally, the fourth node focuses on 'OpenID Connect provider endpoints and custom headers', rather than 'enabling user accounts upon first login'.

ContextualRecall: The score is 0.40 because while the 4th node in the retrieval context supports the mention of 'xwiki.properties' and the property 'oidc.enableUser', it lacks specific instructions on updating the property, saving the file, restarting XWiki, and the automatic activation of new user accounts upon first login.

CustomContextualRelevancy: The retrieval context does not contain any information about automatically activating user accounts with OpenID Connect by modifying the xwiki.properties file or setting the oidc.enableUser property.

Correctness: The actual output correctly identifies the need to modify xwiki.properties to activate users on first login but uses a different property name, 'oidc.activate_user' instead of 'oidc.enableUser'. The explanation includes additional warnings absent in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.5395060255562639

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2370361533375836

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant to the input, addressing the customization of user profile fields in XWiki with OpenID Connect without any irrelevant statements. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node, which discusses the 'xwiki.properties' file crucial for configuring OpenID Connect settings, is ranked higher than all irrelevant

nodes. These irrelevant nodes discuss topics like container authentication and LDAP (second node), general OpenID Connect information without specific field mapping (third node), metadata about the OpenID Connect Authenticator extension (fourth node), and dependencies for the extension (fifth node), none of which directly address the customization of user profile fields.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks information on customizing user profile fields with `oidc.user.mapping` in `xwiki.properties` as described in the expected output.

Correctness: The actual output mentions OpenID Connect and user field mapping but lacks specific instructions found in the expected output, such as editing `xwiki.properties` and using `oidc.user.mapping`.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.35052390756885227

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 0.9090909090909091

ContextualPrecision: 0

ContextualRecall: 0.1111111111111111

CustomContextualRelevancy: 0.0

Correctness: 0.3051636474333159

Reasons:

AnswerRelevancy: The score is 0.78 because the output provides relevant information about setting up group synchronization between XWiki and an OpenID Connect provider, but it includes several irrelevant statements such as source links and references to unrelated documents, which detract from the overall focus on the setup process.

Faithfulness: The score is 0.91 because the actual output references 'Ergebnisdokument Nr. 1' for property listings, which is not mentioned in the retrieval context. This minor discrepancy slightly affects the faithfulness score.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question about setting up group synchronization between XWiki and an OpenID Connect provider. The first node discusses 'authentication management and endpoints for OpenID Connect, but does not provide specific instructions on group synchronization.' The second node provides 'information on container authentication and other authentication methods, but does not address group synchronization between XWiki and OpenID Connect.' The third node describes 'the OpenID Connect Authenticator extension and its capabilities, but does not detail the process of setting up group synchronization.' The fourth node lists 'dependencies for the OpenID Connect Authenticator extension, which is not relevant to the group synchronization setup process.' Finally, the fifth node provides 'general information about the OpenID Connect Authenticator, including compatibility and installation, but does not cover group synchronization setup.'

ContextualRecall: The score is 0.11 because only the mention of `xwiki.properties` in the 1st node in retrieval context supports the expected output, while the rest of the expected output sentences lack corresponding information in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks details on configuring group synchronization, claim names, separators, and mappings as outlined in the expected output.

Correctness: The actual output contains steps for setting up group synchronization but refers to `xwiki.cfg` instead of `xwiki.properties`. It mentions `oidc.syncgroups` while the expected output does not, and lacks detailed configuration examples for `oidc.groups.separator` and `oidc.groups.mapping`.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_018_result.json

Overall_score: 0.605628896072057
Individual_scores:
AnswerRelevancy: 0.7142857142857143
Faithfulness: 1.0
ContextualPrecision: 0.325
ContextualRecall: 0.5
CustomContextualRelevancy: 0.5
Correctness: 0.5944876621466277
Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about attachment restrictions by focusing on relevant aspects, but it includes multiple irrelevant source links that do not directly answer the query about file type and size limitations.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because relevant nodes in the retrieval context are ranked lower than irrelevant nodes. The first node discusses 'trusted domains and PDF export customization', which is not related to attachment limits, and should be ranked lower than the fourth node, which mentions 'Size Limit' and 'Mimetype Restriction', directly relevant to configuring attachment limits and types. Similarly, the second node about 'ODT and PDF export features' should be ranked lower than the fifth node, which provides 'detailed steps on setting the maximum size for attachments'.

ContextualRecall: The score is 0.50 because while the steps for setting the maximum upload size in XWikiPreferences (sentence 2 to 5) are well supported by node 5 in the retrieval context, the initial instructions about configuring MIME types and file size limits (sentence 1 and parts of sentence 2) are not directly supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions configuring attachment size and MIME type restrictions, but lacks specific details on setting MIME types and exact steps for 10 MB limit as in the expected output.

Correctness: The actual output includes the key facts about limiting upload size and MIME types, but it introduces discrepancies in the steps, such as using the Attachment Validation Application instead of modifying xwiki.properties, and provides extra information not present in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_019_result.json

Overall_score: 0.33670319494707285
Individual_scores:
AnswerRelevancy: 0.8181818181818182
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.2020373515006184
Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the question about denying script rights to a space administrator, but it includes irrelevant source links that do not directly address the query.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'validation patterns and programmatic validation' which does not relate to 'rights management or denying script rights'. The second node is about 'template providers and page creation in XWiki', unrelated to 'denying script rights to an area administrator'. The third node describes 'a custom right check configuration in XWiki' but does not address 'denying script rights to an area administrator'.

ContextualRecall: The score is 0.00 because the expected output sentence about rights and admin rights in XWiki is not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about XWiki not supporting the refusal of rights implied by the admin right.

Correctness: Actual output provides a workaround for denying script rights, which contradicts the expected output stating it is not supported in XWiki.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.3154800908901064

Individual_scores:

AnswerRelevancy: 0.5294117647058824

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.125

CustomContextualRelevancy: 0.0

Correctness: 0.2384687806347562

Reasons:

AnswerRelevancy: The score is 0.53 because the output contains several irrelevant statements about database support, notifications, content organization, user creation, and OpenID Connect Authenticator, which do not address the question about configuring team access and editing rights in XWiki. However, some relevant information is likely present, contributing to the score being above zero.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses XWiki's database support and programming features, which are unrelated to configuring access controls for teams. The second node is about the Notifications Application in XWiki, which does not relate to setting access permissions for teams. The third node explains the use of Nested Pages and subwikis, which does not directly address configuring team-specific access rights in XWiki. The fourth node mentions managing users and groups, but does not provide specific instructions on configuring access controls for teams as required by the input. Lastly, the fifth node discusses group synchronization and customization for OpenID Connect, which is not relevant to setting team-specific access permissions in XWiki.

ContextualRecall: The score is 0.12 because only one supportive reason links a sentence in the expected output (sentence 1) to a node in the retrieval context (node 4), while the rest of the sentences lack corresponding information in the nodes, leading to a predominantly unsupportive context.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting wiki-wide or area-specific rights, or inheriting rights for subpages as mentioned in the

expected output.

Correctness: Actual output suggests creating separate (sub)wikis instead of groups for access control. It lacks specific instruction on setting 'View' rights for 'XWikiAllGroup' and 'Edit' rights per team group. Additional information on databases and notifications is irrelevant.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_021_result.json

Overall_score: 0.33668140350982284

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.1867550877256033

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the issue about the notification bell icon, but includes irrelevant source links that do not directly resolve the problem.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job maintaining perfect alignment!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. Each node, from the first to the fifth, fails to address 'enabling notifications in the wiki settings', which is crucial for the input query. Instead, they discuss topics such as 'custom notification filters', 'various notification-related features', 'notification preferences', 'notification application and its features', and 'obtaining a notification RSS feed'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications through the `xwiki.properties` file or setting `notifications.enabled` to `true`, which is the key information in the expected output.

Correctness: The actual output does not mention enabling notifications via `xwiki.properties`, which is a key fact in the expected output. It provides unrelated troubleshooting advice about the notification bell icon and extensions, which are not mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_022_result.json

Overall_score: 0.656549838069111

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Correctness: 0.24456218630940305

Reasons:

AnswerRelevancy: The score is 0.89 because the response mostly addresses the issue of not receiving notifications, but includes irrelevant links that do not directly help solve the problem.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job on maintaining perfect alignment!

ContextualPrecision: The score is 1.00 because all relevant nodes are perfectly ranked higher than irrelevant ones. The first node discusses 'Enable and disable the extension' and 'Enable and disable the email feature', which is crucial for ensuring notifications are activated. The second node includes 'Follow a user', aligning with the input about receiving notifications. The irrelevant nodes, ranked third and fourth, discuss RSS feeds and notification settings without directly addressing enabling notifications or following users, and displaying notification preferences, respectively. Great job on the ranking!

ContextualRecall: The score is 0.50 because while some aspects like enabling notifications and following users are supported by nodes in the retrieval context, specific details such as setting 'notifications.enabled' in 'xwiki.properties', using the 'alert' menu, and the 'network' tab in the user profile are not mentioned in the retrieval context.

CustomContextualRelevancy: Retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' to 'true' and specific steps for following pages or users.

Correctness: The actual output lacks specific instructions about enabling notifications in 'xwiki.properties' and following pages or users, which are critical data points in the expected output. It contradicts by focusing on application and filter settings not mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8657325248893909

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Correctness: 0.6277284826696791

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the question about enabling notifications for one's own actions. However, it includes irrelevant links and additional information that do not directly answer the question, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which states 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' is ranked first and directly addresses the input question. The irrelevant nodes, discussing topics such as RSS feeds, general notification settings, and technical aspects, are appropriately ranked lower.

ContextualRecall: The score is 1.00 because all sentences in the expected output are perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context confirms that by default, users do not receive notifications for their own actions and mentions the option to change this setting. However, it does not explicitly mention the 'Options de filtrage avancées' section.

Correctness: The actual output describes how to enable notifications for one's own actions, which aligns with the expected output about disabling a filter to receive such notifications. However, it includes

procedural steps and additional external links not present in the expected output, and lacks direct mention of the 'Options de filtrage avancées' section.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.822027349428013

Individual_scores:

AnswerRelevancy: 0.6923076923076923

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Correctness: 0.6898564042603862

Reasons:

AnswerRelevancy: The score is 0.69 because the answer partially addresses the question about disabling notifications, but it includes multiple irrelevant source links that do not directly contribute to solving the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant ones. The first three nodes effectively address the input by discussing 'Choosing the default auto-watched mode', 'Notifications Auto Watch Preferences', and 'Auto Watch' and 'Filters', which are all pertinent to managing notification settings. The irrelevant node, which discusses 'RSS feeds and general notification settings', is appropriately ranked last, ensuring a perfect score. Great job!

ContextualRecall: The score is 0.75 because while most sentences in the expected output are supported by nodes in the retrieval context, such as node 1 and node 2, the sentence about being notified of changes on your work lacks direct support from the retrieval context.

CustomContextualRelevancy: The retrieval context includes information about auto-watched pages and custom notification filters, aligning with the expected output. However, it lacks explicit mention of modifying these settings in the user's profile notification settings.

Correctness: Both outputs mention 'Auto Watch' and modifying notification settings. The actual output adds steps to disable 'Auto Watch' and provides source links, which are not in the expected output.

There's no mention of 'Filtres personnalisés' in the actual output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.369055137867685

Individual_scores:

AnswerRelevancy: 0.625

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.18933082720611033

Reasons:

AnswerRelevancy: The score is 0.62 because the output partially addresses the configuration issue related to OpenID Connect in XWiki, but it includes multiple irrelevant source links that do not directly contribute to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based authentication and session cookies', which is unrelated to enabling user accounts automatically on first login using OpenID Connect. The second node talks about 'container authentication and LDAP', missing the specific property `oidc.enableUser` in `xwiki.properties` needed for automatic account activation. The third node, despite mentioning OpenID Connect, does not provide information about configuring `oidc.enableUser` to automatically activate user accounts. The fourth node lists 'dependencies for the OpenID Connect extension', but does not address user account activation. The fifth node is about 'LDAP and Active Directory integration', which is not relevant to the configuration needed.

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention 'xwiki.properties' and OpenID Connect, they do not provide specific details about adding or updating properties, saving the file, or restarting XWiki, which are crucial for the expected output sentences 2, 3, and 5.

CustomContextualRelevancy: The retrieval context does not contain information about adjusting the xwiki.properties file or setting oidc.enableUser to true for automatic user activation via OpenID Connect.

Correctness: The actual output references the wrong file (`xwiki.cfg` instead of `xwiki.properties`) and the wrong property (`xwiki.authentication.createuser` instead of `oidc.enableUser`). It suggests creating users rather than activating them. It does mention restarting XWiki, which aligns with the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.4795238996195658

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.0

Correctness: 0.4200005405745374

Reasons:

AnswerRelevancy: The score is 0.86 because the output mostly addresses the question about customizing user profile fields in XWiki, but it includes irrelevant source links that do not directly contribute to answering the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect endpoints,' but does not mention 'user profile field customization or mapping.' The second node is about 'container authentication and various authentication methods like LDAP and Active Directory,' without any reference to 'customizing user profile fields or mapping them.' The third node provides 'general information about the OpenID Connect Authenticator, including installation and compatibility,' but does not address 'user profile field customization or mapping.' The fourth node lists 'dependencies for the

OpenID Connect Authenticator extension,' which is unrelated to 'user profile field customization or mapping.' Lastly, the fifth node describes 'configuration guides for specific systems and general configuration,' but does not mention 'mapping user profile fields or customizing them.'

ContextualRecall: The score is 0.60 because while the nodes in retrieval context support the general concept of configuring OpenID Connect and mapping user profile fields in 'xwiki.properties', they lack specific examples and instructions found in the expected output.

CustomContextualRelevancy: The retrieval context does not mention customizing user profile fields or mapping them in the xwiki.properties file using oidc.user.mapping, which is the core information in the expected output.

Correctness: The actual output uses `xwiki.cfg` instead of `xwiki.properties` and provides a different mapping method using JSONPath. It lacks the specific example from the expected output and introduces new mapping elements not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.535672508761512

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.3076923076923077

CustomContextualRelevancy: 0.3

Correctness: 0.2730094115434303

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider, with no irrelevant information included. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which 'directly addresses group synchronization by mentioning the `oidc.groups.mapping` property and the process of synchronizing group claims', is ranked third. Irrelevant nodes, such as the first node that 'discusses authentication management and OpenID Connect endpoints, but does not mention group synchronization', and the second node that 'provides information on different authentication methods and extensions, but does not address group synchronization with OpenID Connect', are ranked higher than the relevant node.

ContextualRecall: The score is 0.31 because while the 3rd node in the retrieval context supports the general concept of configuring group synchronization in XWiki, it lacks specific details such as locating and editing the 'xwiki.properties' file, providing examples, and instructions on saving and restarting XWiki, which are present in the expected output.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping but lacks specific details on configuring `oidc.groups.claim`, `oidc.groups.separator`, and saving changes in `xwiki.properties` as outlined in the expected output.

Correctness: Actual output uses `xwiki_groups` in `oidc.userinfoclaims` instead of `oidc.groups.claim`. It also lacks steps on `oidc.groups.separator`, sequence of events does not fully align, and it suggests additional customization options not in expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3181364330516399

Individual_scores:

AnswerRelevancy: 0.7058823529411765

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.20293624536866267

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about file upload restrictions, but it includes multiple irrelevant source links that do not directly relate to the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'language support and translations in XWiki', which is unrelated to configuring file upload restrictions. The second node is about 'creating tour guides and administrating wiki features', which does not relate to setting file upload restrictions. The third node discusses 'CSS and XSL for PDF exports', which is not relevant to configuring file upload restrictions. The fourth node is about 'various features and macros in XWiki', none of which pertain to file upload restrictions. The fifth and sixth nodes contain 'image links and references', which are not relevant to configuring file upload restrictions.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, as the context lacks any relevant information or instructions related to the topics covered in the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information related to configuring MIME types or setting maximum attachment size in XWiki, as described in the expected output.

Correctness: The actual output fails to mention editing the xwiki.properties file and configuring MIME types specifically. It lacks precise instructions for setting the upload size in bytes, only providing a general approach. The expected output gives detailed steps, which are missing in the actual output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.6124564545340901

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.2247387272045409

Reasons:

AnswerRelevancy: The score is 0.75 because the answer addresses the question about refusing script rights to a space administrator, but includes irrelevant source links that do not directly contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because the second node in the retrieval context is relevant as it discusses the 'Script' right and its default permissions, which aligns with the input query. However, the first node is irrelevant as it discusses language preferences and translations in XWiki, which is unrelated to script rights or administrative permissions. Additionally, the third node is also irrelevant as it lists various rights without addressing the specific issue of refusing script rights to an administrator. These irrelevant nodes are ranked higher than the relevant node, affecting the score.

ContextualRecall: The score is 1.00 because everything aligns perfectly! Great job on ensuring all the necessary information is covered accurately.

CustomContextualRelevancy: The retrieval context mentions that certain rights are not deniable, but does not explicitly state that admin rights cannot be refused as in the expected output.

Correctness: The actual output provides detailed steps for refusing script rights, which contradicts the expected output stating it cannot be done. It lacks coherence with the expected key fact that refusing administrator rights is not supported in XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.3682516154894054

Individual_scores:

AnswerRelevancy: 0.7647058823529411

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.4448038105834914

Reasons:

AnswerRelevancy: The score is 0.76 because the response partially addresses the question about configuring team spaces in XWiki, but includes multiple irrelevant statements that do not directly answer the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'the structure of a page in XWiki' which does not provide information on setting access rights for teams. The second node talks about 'sending representations and modifying resources', unrelated to configuring access rights for teams in XWiki. The third node is about 'editing and managing comments', which does not relate to configuring access rights for team spaces in XWiki. The fourth node explains 'the concept of spaces and pages in XWiki', but lacks guidance on setting access permissions for teams. The fifth node discusses 'actions needed after running a migrator', not addressing access rights for team spaces. Lastly, the sixth node is about 'configuration display settings in XWiki', unrelated to setting access permissions for team spaces.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain information about creating groups, defining rights at wiki or space levels, or inheriting rights for child pages as mentioned in the expected output.

Correctness: The actual output describes a different method for setting permissions using 'Rôles' which is not mentioned in the expected output. Both outputs ensure visibility for all teams and modification rights for individual teams, but the specific steps and details differ significantly.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_031_result.json

Overall_score: 0.6775766226412651
Individual_scores:
AnswerRelevancy: 0.6363636363636364
Faithfulness: 1.0
ContextualPrecision: 0.5
ContextualRecall: 0.7142857142857143
CustomContextualRelevancy: 0.7
Correctness: 0.51481038519824
Reasons:

AnswerRelevancy: The score is 0.64 because the response partially addresses the question about authenticating users with access tokens, but includes multiple irrelevant source links that do not contribute to answering the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because relevant nodes in the retrieval context are not consistently ranked higher than irrelevant nodes. The first node, ranked 1, is irrelevant as it focuses on 'LDAP configuration' without mentioning access tokens. The second node, ranked 2, is relevant as it discusses 'configuration of authorized applications', which aligns with using access tokens for authentication. The third node, ranked 3, is irrelevant since it addresses 'OpenID Connect scopes and response types', not directly related to access tokens. The fourth node, ranked 4, is relevant because it describes 'authentication with JWT', supporting the use of access tokens. The fifth node, ranked 5, is irrelevant as it covers 'various authentication mechanisms in XWiki' without focusing on token-based authentication.

ContextualRecall: The score is 0.71 because while several key aspects like authenticating users with JWT, enabling the token authenticator, and configuring authorized applications are well-supported by nodes in the retrieval context, specific details about generating a JWT token with required claims and including the token in the 'Authorization' header are missing.

CustomContextualRelevancy: The retrieval context mentions the use of JWT for authentication and the configuration of authorized applications, but lacks details on token claims and the specific process of including the token in the request header.

Correctness: The actual output matches the expected output in terms of JWT token authentication and property settings but lacks details on claims, signing key, and 'Authorized Applications' configuration. It also incorrectly mentions a fallback authenticator, which is not in the expected output.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_032_result.json

Overall_score: 0.31684636081682344
Individual_scores:
AnswerRelevancy: 0.8666666666666667
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0

Correctness: 0.03441149823427401

Reasons:

AnswerRelevancy: The score is 0.87 because the response mostly addresses the question about authenticating users with access tokens, but it includes source links that are not directly related to the topic, slightly reducing its relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'model validation strategies' without mentioning authentication or access tokens. The second node covers 'attachment validation and configuration', unrelated to user authentication with access tokens. The third node provides 'information about the admin UI for model validation', which does not pertain to authenticating users with access tokens.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT as mentioned in the expected answer.

Correctness: The actual output provides general information about token-based authentication but lacks specific details from the expected output, such as JWT requirements, configuration steps, and limitations of the authenticator.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.308138737976724

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.13454671357462955

Reasons:

AnswerRelevancy: The score is 0.71 because the response addresses the question about authenticating users with access tokens, but includes irrelevant links that do not directly pertain to the topic, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'upgrading XWiki' and does not relate to 'user authentication or token-based authentication.' The second node talks about 'features of the Administration Application, such as Users & Rights,' without mentioning 'token-based authentication.' The third node is about 'managing rights and user profiles,' again with no mention of 'token-based authentication methods.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be linked to any nodes in the retrieval context. The retrieval context lacks information on authenticating users with access tokens, JWTs, or any related configurations and limitations.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are key elements in the expected output.

Correctness: The actual output fails to align with the expected output as it does not mention the JWTTokenAuth extension, the specific requirements for JWT claims, or the steps for enabling the authenticator. It only provides a general idea about JWT usage in web applications.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

Model: AI.Models.qa_phi3_medium-128k_14b_Q4

File: qa_001_result.json

Overall_score: 0.3009989903656263

Individual_scores:

AnswerRelevancy: 0.8125

Faithfulness: 0.8

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.19349394219375757

Reasons:

AnswerRelevancy: The score is 0.81 because the response largely addresses the issue of the bell icon not being available in XWiki, but includes irrelevant source links and a statement about offline extension installation, which do not directly help resolve the specific problem.

Faithfulness: The score is 0.80 because the actual output implies that installing extensions offline might be possible through complex manual methods, whereas the retrieval context clearly states that offline installation is not supported.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'installation instructions and prerequisites', the second node describes 'notification features and appearance', the third node lists 'dependencies', and the fourth node talks about 'customizing notification templates'. None of these address the issue of enabling notifications in the wiki, which is the focus of the input.

ContextualRecall: The score is 0.00 because the retrieval context does not provide any information related to enabling notifications in the wiki or the `xwiki.properties` file, which are crucial for the expected output.

CustomContextualRelevancy: The retrieval context lacks any information about enabling notifications through the `xwiki.properties` file, which is the core instruction in the expected output.

Correctness: The actual output provides various troubleshooting steps but lacks specific instructions on enabling notifications via `xwiki.properties`, as outlined in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.5656465255342522

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 0.42857142857142855

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Correctness: 0.4153077246340847

Reasons:

AnswerRelevancy: The score is 0.80 because the response largely addresses the issue of not receiving notifications in XWiki, but it includes irrelevant source links that do not directly contribute to

solving the problem.

Faithfulness: The score is 0.43 because the actual output contains several contradictions: it incorrectly claims that nothing is watched by default, whereas earlier versions of XWiki had users watching the whole wiki by default; it states that no notifications are received until manual actions are taken, contrary to the fact that application types are enabled by default in later versions; it suggests automatic enabling of application types upon certain actions, which is not supported by the context; and it misrepresents the purpose of the 'network' tab in user profiles.

ContextualPrecision: The score is 0.75 because the first node in the retrieval context is relevant, explaining how to follow a user to receive notifications, and is ranked appropriately. However, the second and third nodes, which discuss filters, settings, and RSS feeds without addressing enabling notifications or subscribing to specific content, are ranked higher than the fourth node, which is relevant as it outlines ways to receive notifications through the Notifications Application. The fifth node, discussing the 'What's New' application, is also irrelevant and should be ranked lower.

ContextualRecall: The score is 0.50 because while the retrieval context supports details about enabling notifications through the 'alert' menu, following users, and checking the 'network' tab (as mentioned in sentences 4, 5, and 6 of the expected output), it lacks information on enabling notifications via `xwiki.properties` and subscribing to pages or users (sentences 1, 2, and 3).

CustomContextualRelevancy: The retrieval context includes details on following users and notification settings, but lacks explicit mention of enabling notifications via 'xwiki.properties' and watching pages or wikis.

Correctness: The actual output omits enabling notifications in xwiki.properties and subscribing to pages or users. It mentions watching users and application types but lacks details on enabling notifications through the alert menu. Some filtering options are covered, but the focus is not aligned with the expected output instructions.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9004205638631171

Individual_scores:

AnswerRelevancy: 0.7272727272727273

Faithfulness: 0.8

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.8752506559059755

Reasons:

AnswerRelevancy: The score is 0.73 because the response partially addresses the question about enabling notifications for one's own actions in XWiki, but includes irrelevant source links that do not directly answer the query.

Faithfulness: The score is 0.80 because the actual output incorrectly states that the path to modify notification settings in XWiki is Administration > Notifications, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant ones. Great job on getting the order spot on!

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the nodes in the retrieval context, demonstrating perfect alignment.

CustomContextualRelevancy: The retrieval context contains all the information found in the expected answer, including the default behavior of not receiving notifications about one's own activity unless explicitly targeted, and the ability to disable this filter in the 'Advanced filtering options' section of the

notification settings.

Correctness: The Actual Output aligns with the Expected Output regarding default notification settings and user ability to change them. It provides additional clarity on how to modify these settings, which doesn't contradict the Expected Output. The mention of XWiki version 14.1+ is extra but not a factual inaccuracy.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.3925233142335871

Individual_scores:

AnswerRelevancy: 0.35294117647058826

Faithfulness: 0.7142857142857143

ContextualPrecision: 0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Correctness: 0.2379129946452201

Reasons:

AnswerRelevancy: The score is 0.35 because the output contains multiple irrelevant statements that do not directly address how to disable notifications in XWiki, such as providing a source link, unrelated navigation suggestions, and discussing unrelated features. However, the score is not lower because there may be some relevant information amidst the irrelevant content.

Faithfulness: The score is 0.71 because the actual output incorrectly states that all application types are enabled by default for new users in version 15.5+, while the retrieval context indicates that nothing is watched by default. Additionally, the actual output mentions an 'Advanced filtering options' section for controlling notifications about one's own activity, which is not found in the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question about disabling notifications in XWiki. For example, the first node discusses 'getting a notification RSS feed and settings for notifications', but does not mention disabling notifications or automatic page watching. Similarly, the second node talks about 'following a user and default watch settings', which is not related to disabling notifications. All nodes fail to address the specific query, leading to a score of 0.00.

ContextualRecall: The score is 0.75 because the retrieval context supports the expected output regarding automatic page watching and notification settings, as seen in nodes 2 and 5. However, it lacks explicit information about deleting existing notification filters, which affects the completeness of the match.

CustomContextualRelevancy: The retrieval context mentions default notification settings and user profile adjustments but lacks specific details about automatic page watching and deleting notification filters.

Correctness: The actual output describes notification settings and user control but lacks specifics about automatic page watching after major modifications and deleting filters from the Custom Filters list, which are key aspects of the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8116396216730285

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Correctness: 0.4126948728953138

Reasons:

AnswerRelevancy: The score is 0.86 because the output mostly addresses the question about configuring OpenID Connect in XWiki, but it includes an irrelevant statement about LDAP Authenticator, which is unrelated to the user's query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions the property 'oidc.enableUser=true', is ranked first. This directly addresses the configuration needed to enable user accounts on first login using OpenID Connect. The irrelevant nodes, discussing topics like domain-based authentication, general OpenID Connect information, container authentication, and LDAP configuration, are all ranked lower, ensuring precise retrieval.

ContextualRecall: The score is 0.60 because while the retrieval context accurately captures the key details about adjusting the 'xwiki.properties' file and setting the 'oidc.enableUser' property (as seen in nodes related to sentences 1, 2, and 3), it lacks information on saving the file, restarting the XWiki instance, and the automatic enabling of new user accounts upon first login, which are crucial for a complete understanding.

CustomContextualRelevancy: The retrieval context includes the exact property '#oidc.enableUser=true' that matches the expected answer, indicating the necessary configuration change in the xwiki.properties file.

Correctness: The actual output suggests modifying 'xwiki.cfg' instead of 'xwiki.properties' and lacks instruction to save and restart XWiki, but it correctly identifies setting 'oidc.enableUser' to true.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.9220803778494632

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.8

CustomContextualRelevancy: 1.0

Correctness: 0.8324822670967791

Reasons:

AnswerRelevancy: The score is 0.90 because the answer is mostly relevant and provides useful information about customizing user profile fields in XWiki. However, it includes a source link that does not directly address the specific question about mapping additional fields, which slightly detracts from its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job maintaining accuracy!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user' and provides an example of 'oidc.user.mapping=myxproperty1=\${oidc.user.subject}', is ranked first. This directly relates to the input

about mapping additional user profile fields. All irrelevant nodes, discussing topics like access tokens and scopes, endpoints and HTTP methods, group synchronization, and in-wiki configuration, are ranked lower, ensuring perfect contextual precision.

ContextualRecall: The score is 0.80 because the retrieval context nodes provide detailed support for mapping user profile fields using 'oidc.user.mapping', as described in sentences 1 to 4 of the expected output. However, it lacks information about saving the file and restarting XWiki, which affects the completeness of the context.

CustomContextualRelevancy: The retrieval context contains the information about customizing user profile fields using the 'oidc.user.mapping' property in the 'xwiki.properties' file, matching the expected answer.

Correctness: The actual output correctly explains mapping user profile fields using oidc.user.mapping in xwiki.properties, aligning with the expected output. However, it introduces additional details about file location and administrative access that are not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8073287990052678

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8571428571428571

ContextualPrecision: 1.0

ContextualRecall: 0.4166666666666667

CustomContextualRelevancy: 1.0

Correctness: 0.5701632702220831

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider. Great job!

Faithfulness: The score is 0.86 because the actual output incorrectly states that the claim to indicate the groups a user belongs to is 'oidc.groups.claim', while the retrieval context specifies it as 'xwiki_groups'.

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first three nodes in the retrieval context provide detailed information on setting properties like 'oidc.groups.claim', 'oidc.groups.mapping', and 'oidc.groups.separator', which are directly relevant to configuring group synchronization. The irrelevant nodes, ranked fourth and fifth, focus on general OpenID Connect information and authentication management, which do not address the specific group synchronization process.

ContextualRecall: The score is 0.42 because while some specific configuration instructions like setting 'oidc.groups.claim' and 'oidc.groups.separator' are supported by nodes in the retrieval context, many general instructions and introductory statements in the expected output are not directly found in the retrieval context.

CustomContextualRelevancy: The retrieval context contains all necessary details for configuring group synchronization, including setting the 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' properties, as well as saving the 'xwiki.properties' file, matching the expected output.

Correctness: Both outputs cover configuring 'oidc.groups.claim' and 'oidc.groups.mapping', but the actual output mentions 'oidc.userinfoclaims', which is not in the expected output. Actual output includes some key steps but lacks specific instructions about editing 'xwiki.properties' directly and restarting XWiki, which are present in expected output.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_008_result.json

Overall_score: 0.6530171595457864
Individual_scores:
AnswerRelevancy: 0.9285714285714286
Faithfulness: 1.0
ContextualPrecision: 0.3333333333333333
ContextualRecall: 0.72727272727273
CustomContextualRelevancy: 0.3
Correctness: 0.6289254680972296
Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to the question about limiting attachment types and sizes, but it includes a general link to XWiki basic concepts, which is not directly relevant to the specific query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which provides information about setting the maximum size of an attachment and mentions configuration steps in XWikiPreferences, is ranked third. However, irrelevant nodes, such as the first node discussing linking to external files and image display, and the second node explaining the structure of XWiki pages, are ranked higher than the relevant node. This misordering affects the score, as the relevant information is not prioritized.

ContextualRecall: The score is 0.73 because while many steps in the expected output are supported by nodes in the retrieval context, such as node 3 supporting the configuration steps for maximum attachment size, some introductory and specific configuration details are missing from the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific details on configuring 'xwiki.properties' and setting size to 10MB.

Correctness: Actual output correctly explains setting the maximum attachment size and mentions mime type configuration, but incorrectly suggests using `attachment.download.forceDownload` for mime-types instead of `attachment.upload.allowList` in `xwiki.properties`.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_009_result.json

Overall_score: 0.6318548193996071
Individual_scores:
AnswerRelevancy: 0.8125
Faithfulness: 0.8333333333333334
ContextualPrecision: 0
ContextualRecall: 1.0
CustomContextualRelevancy: 1.0
Correctness: 0.14529558306430865
Reasons:

AnswerRelevancy: The score is 0.81 because the main content effectively addresses the question about denying script rights to a space administrator, but it is slightly lowered due to the inclusion of multiple source links that do not directly pertain to the input question.

Faithfulness: The score is 0.83 because the actual output incorrectly claims that script right is allowed for all users at the main wiki level by default, while the retrieval context specifies that in XWiki 14.10+, script right is not given by default to all users at the main wiki level.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses programming rights and script execution but does not address denying script rights to a space administrator. The second node lists various rights and their properties but does not mention the inability to deny script rights to a space administrator. The third node explains the script right and its default status but does not mention the inability to deny it to a space administrator. The fourth node provides code snippets related to checking access rights but does not discuss the specific issue of denying script rights to a space administrator. The fifth node is about converting rights during page migration and does not relate to denying script rights to a space administrator.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the nodes in the retrieval context, indicating perfect alignment.

CustomContextualRelevancy: The expected answer states that rights implied by admin right cannot be denied, which is supported by the retrieval context indicating that admin rights are not deniable.

Correctness: The actual output provides instructions on how to deny script rights to a space admin, contradicting the expected output that states it is not supported in XWiki. The actual output includes additional steps and information not present in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7138053672715047

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 0.875

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.875

CustomContextualRelevancy: 0.5

Correctness: 0.46616553696236185

Reasons:

AnswerRelevancy: The score is 0.90 because the response mostly addresses how to configure access rights in XWiki, but it includes irrelevant source links that do not directly contribute to solving the problem.

Faithfulness: The score is 0.88 because the actual output incorrectly claims that an explicit denial does not block inheritance, while the retrieval context clarifies that when a right is allowed at a given level, it is implicitly denied to others at the same level.

ContextualPrecision: The score is 0.67 because relevant nodes in the retrieval context are ranked higher than some irrelevant nodes, but not all. The first node discusses 'rights inheritance and administration features', which is relevant, and is correctly ranked higher. However, the second node, which focuses on 'the parent/child relationship and the introduction of nested pages', is irrelevant and should be ranked lower than the fourth node, which provides information on 'XWiki's permission system, including setting view and edit rights'. Additionally, the third node about 'string formatting and user identification' is also irrelevant and should be ranked lower than the sixth node, which explains 'basic rules for setting rights in XWiki'.

ContextualRecall: The score is 0.88 because most of the expected output sentences are well-supported by nodes in the retrieval context, such as creating groups, setting wiki-wide and

space-level rights, and inheriting rights for child pages. However, there is a lack of specific mention in the retrieval context about adjusting rights for additional teams or spaces, which slightly affects the score.

CustomContextualRelevancy: The retrieval context mentions rights inheritance and setting rights at different levels, which aligns with setting space-level rights and inheriting rights for child pages. However, it lacks specific details about creating groups for each team and setting the 'View' right for 'XWikiAllGroup'.

Correctness: The actual output partially aligns with the expected output by setting 'View' rights at the wiki level and restricting edit rights to respective teams, but it includes unnecessary administrator rights details and lacks specific steps for group creation and child page inheritance.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.45411768630989485

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.17568651001623234

Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses the issue of the unavailable bell icon, but it includes irrelevant source links that do not directly contribute to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node in retrieval context, which discusses 'editing the `xwiki.properties` file' to enable notifications, is ranked last (6th node). The irrelevant nodes, such as the first node that 'discusses custom notification filters' without addressing enabling notifications, are ranked higher, reducing the score.

ContextualRecall: The score is 0.50 because while the expected output mentions enabling notifications, which is supported by the 5th node in the retrieval context, it lacks specific details about setting `notifications.enabled` in `xwiki.properties` to `true`, which is not found in any node in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the `xwiki.properties` file or setting `notifications.enabled` to `true`.

Correctness: The actual output addresses potential issues with the notifications feature but fails to mention enabling notifications via `notifications.enabled` in `xwiki.properties`. It includes unrelated troubleshooting steps not present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3395941898643187

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.14867625029702353

Reasons:

AnswerRelevancy: The score is 0.89 because the response mostly addresses the issue of not receiving notifications, but includes some irrelevant parts like a source link, a console command suggestion, and a closing statement that do not directly help with the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node contains 'a list of image links and does not provide any information related to notifications or their settings', the second node 'contains image links and references to various projects, which do not relate to the topic of enabling notifications', and the third node is 'a series of image links and references, offering no relevant information about notification settings or issues'. These irrelevant nodes should be ranked lower than any relevant nodes.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context contains image links and references that do not match any information in the expected output about enabling notifications or following pages and users.

Correctness: The actual output lacks specific instructions like setting `notifications.enabled` in `xwiki.properties` to `true` and following pages or users. It introduces unrelated steps like checking 'Admin-Verzeichnis' and using a 'Konsolenbefehl'.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.27648800589947636

Individual_scores:

AnswerRelevancy: 0.30434782608695654

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.35458020930990186

Reasons:

AnswerRelevancy: The score is 0.30 because the output contains multiple irrelevant statements that do not address the question about enabling notifications for one's own actions. The statements focus on notifications about others' actions and settings unrelated to the user's query, which significantly lowers the relevancy score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customizing page templates, internationalization, accessibility, and performance', which is unrelated to notification settings. The second node is about 'an application for administrating wiki features', not addressing notification settings or user activity. The third node covers 'various extensions such as Office Importer, Font Awesome Icon Theme', none of which relate to user notification settings. Therefore, no relevant nodes are ranked higher than irrelevant ones.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be linked to any nodes in the retrieval context, as key details about notification settings and filter options are missing.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters, which are key points in the expected answer.

Correctness: The actual output provides detailed notification settings but lacks mention of disabling filters in 'Erweiterte Filteroptionen' as stated in the expected output. It also includes additional information not present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.2757011913607686

Individual_scores:

AnswerRelevancy: 0.6521739130434783

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.0020332351211331348

Reasons:

AnswerRelevancy: The score is 0.65 because the output contains several irrelevant statements about unrelated topics like attachments, XWiki syntax, and security modules, which do not directly address the question about disabling notifications. However, the score is not lower because there might still be some relevant information regarding the actual question.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job on maintaining perfect alignment!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files and related features', which is not about managing notifications. The second node talks about 'moving attachments and related actions', again unrelated to notification settings. The third node provides a 'list of macros and their functions', which does not pertain to disabling notifications. The fourth node describes an 'application for administrating wiki features', not involving notification management. The fifth node lists 'various extensions and their functionalities', none of which relate to notification settings. Finally, the sixth node lists 'various applications and their purposes', but none are about notification management.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be matched to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to automatic page following or notification settings as described in the expected output.

Correctness: The actual output discusses feedreader settings and steps unrelated to the expected output about automatic page tracking and notification settings.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.28162606575512544

Individual_scores:

AnswerRelevancy: 0.6923076923076923

Faithfulness: 0.5833333333333334

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.014115368889727128

Reasons:

AnswerRelevancy: The score is 0.69 because the response partially addresses the configuration of automatic user activation with OpenID Connect, but includes several irrelevant statements that detract from directly answering the question.

Faithfulness: The score is 0.58 because the actual output contains several inaccuracies: it incorrectly states that the xwiki.cfg file needs configuration for OpenID Connect, that the OIDC authenticator appends a certain property, and that setting a session cookie on a top-level domain covers all wikis. It also wrongly claims that setting oidc.skipped to true bypasses OpenID Connect and mentions an incorrect rating for the authenticator.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based authentication and session cookie configuration', which is unrelated to the question about automatic user activation. The second node provides 'general information about OpenID Connect authentication and installation instructions', but lacks specifics on automatic user activation. The third node focuses on 'OpenID Connect provider and general authentication capabilities', without addressing the needed configuration. Finally, the fourth node lists 'various OpenID Connect endpoints and configuration options', but does not include information about enabling automatic user activation.

ContextualRecall: The score is 0.40 because while the retrieval context mentions modifying the `xwiki.properties` file, it lacks specific details about the `oidc.enableUser` property, the need to restart the XWiki instance, and the automatic activation of new user accounts upon first login via OpenID Connect.

CustomContextualRelevancy: The retrieval context does not mention the `oidc.enableUser` property or the process of updating `xwiki.properties` to automatically activate user accounts on first login with OpenID Connect.

Correctness: Actual output discusses configuration in xwiki.cfg and unrelated topics like domain configuration and unrelated installation instructions, missing key steps about modifying xwiki.properties and setting oidc.enableUser to true.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.33984716388416825

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.75

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2890829833050094

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and directly addresses the input without any irrelevant statements. Great job!

Faithfulness: The score is 0.75 because the actual output incorrectly asserts that the field `**oidc.scope.openid**` is required for authentication, while the retrieval context only lists 'openid' as a standard scope without marking it as required. Additionally, the actual output inaccurately describes a

'User Configuration' menu for specifying custom headers for non-OIDC providers, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management class and OpenID Connect endpoints', the second node talks about 'authentication methods including OpenID Connect', the third node provides a 'general description of the OpenID Connect Authenticator', the fourth node describes 'compatibility and installation details', and the fifth node lists 'dependencies for the OpenID Connect Authenticator extension'. None of these nodes address user profile field customization or the `oidc.user.mapping` property, which are crucial for the input query.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output could be linked to any nodes in the retrieval context, as the content regarding customizing user profile fields, specific instructions, examples, and configuration steps are absent in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks information on customizing user profile fields using `oidc.user.mapping` in `xwiki.properties`, which is detailed in the expected output.

Correctness: The actual output lacks specific details on how to map user profile fields in `xwiki.properties` using `oidc.user.mapping` and includes additional irrelevant steps like installing the OIDC authenticator, which are not in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.3499586096223295

Individual_scores:

AnswerRelevancy: 0.92

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.1797516577339772

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant to the question about setting up group synchronization between XWiki and an OpenID Connect provider. However, it includes source links that do not directly address the setup process, slightly reducing the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'general OpenID Connect provider configuration and endpoints' but does not address group synchronization, which is crucial for the input query. The second node is about 'container authentication' and lacks details on group synchronization. The third node mentions 'OpenID Connect authentication and synchronization' but does not provide specific instructions for group synchronization. The fourth node lists 'dependencies for the OpenID Connect extension' without relating to group synchronization. Lastly, the fifth node provides 'general information about the OpenID Connect authenticator' but lacks group synchronization details.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, as there are no supportive reasons and multiple unsupportive reasons for each sentence.

CustomContextualRelevancy: The retrieval context lacks specific instructions on configuring group synchronization properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' as outlined in the expected output.

Correctness: The actual output focuses on installing and configuring the OpenID Connect Authenticator, whereas the expected output details specific properties in xwiki.properties for group synchronization. Key points like setting oidc.groups.claim, separator, and mapping are missing in actual output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6180074581197905

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0.325

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Correctness: 0.583044748718743

Reasons:

AnswerRelevancy: The score is 0.80 because the response mostly addresses the question about setting upload restrictions for attachments, but it includes irrelevant links about exporting features, XWiki syntax, and document lifecycle features, which are not directly related to the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the first three nodes in the retrieval context are irrelevant, discussing 'trusted domains and PDF export customization', 'ODT and PDF export features', and 'various macros and features', which do not pertain to configuring upload restrictions for attachments. The relevant nodes, which discuss 'Size Limit' and 'Mimetype Restriction', and provide 'detailed steps on setting the maximum size for attachments', are ranked fourth and fifth, lowering the score. However, the presence of these relevant nodes contributes to the score being above zero.

ContextualRecall: The score is 0.50 because while the steps to change the maximum upload size in XWikiPreferences (sentences 2 to 6) are well-supported by the 5th node in the retrieval context, the initial instructions regarding MIME type configuration (sentences 1 and 7 to 11) are not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context partially covers the expected answer by mentioning the configuration of maximum upload size and MIME type restrictions, but lacks specific details such as the exact MIME type configuration line and the step-by-step process for setting the maximum size to 10 MB.

Correctness: Actual output correctly details setting the maximum upload size to 10MB and provides related steps, but lacks specific configuration of MIME types in 'xwiki.properties' as shown in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_019_result.json

Overall_score: 0.29260924697552043

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.00565548185312246

Reasons:

AnswerRelevancy: The score is 0.75 because the answer partially addresses the question about denying script rights to a domain administrator, but includes multiple irrelevant statements such as source links and image descriptions that do not directly contribute to solving the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'JavaScript validation and field validation rules', which are unrelated to script rights or administrative permissions in XWiki. The second node is about 'templates and page creation in XWiki', which does not relate to administrative rights or script permissions. The third node involves a 'custom right check class in XWiki', but it does not address the specific issue of denying script rights to a space administrator. Therefore, none of the nodes provide relevant information, leading to a score of 0.00.

ContextualRecall: The score is 0.00 because the expected output sentence about rights management in XWiki cannot be linked to any information in the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about admin rights or their implications, which is the focus of the expected answer.

Correctness: The actual output introduces a detailed procedure for preventing script execution, which is not present in the expected output. The expected output simply states that rights implied by the admin right cannot be denied, with no mention of scripts or procedures.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.326883240494203

Individual_scores:

AnswerRelevancy: 0.6875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.27379944296521763

Reasons:

AnswerRelevancy: The score is 0.69 because the response partially addresses the question about configuring team access in XWiki, but it includes multiple irrelevant source links that do not directly relate to the task, which lowers the relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to configuring access control for teams in XWiki. The first node discusses 'XWiki's database support and programming features', which are unrelated to the input query. The second node focuses on 'notifications and the XWiki Development Team', which does not pertain to the task at hand. The third node covers 'using nested pages and subwikis', failing to address the specific need for configuring access rights for teams. The fourth node mentions 'managing users and groups' but lacks specific information on team access configuration. Lastly, the fifth node discusses 'group synchronization and customization', which is not directly related to the input query.

ContextualRecall: The score is 0.00 because none of the expected output sentences can be linked to any nodes in the retrieval context, indicating a complete lack of relevant information.

CustomContextualRelevancy: The retrieval context lacks any information related to setting access controls, creating groups, or managing rights in XWiki as described in the expected output.

Correctness: The actual output suggests creating separate (Sub)-Wikis instead of groups, and using ACLs and 'IstAdmin' group, which deviates significantly from the expected output's steps like setting wiki-wide and area-level rights. Concepts like 'Sider' and 'Category' do not align with the expected approach. Key discrepancies in approach and terminology exist.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.43115784432895193

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.2869470659737118

Reasons:

AnswerRelevancy: The score is 0.80 because the response partially addresses the issue of the unavailable bell icon, but includes irrelevant information such as source links and unrelated potential causes, which detracts from its overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output perfectly aligns with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'custom notification filters and their behavior', but does not address enabling notifications in the wiki settings, which is the core issue in the input. Similarly, the second node lists 'various links related to notification settings and preferences', yet fails to mention enabling notifications in the wiki settings. This pattern continues with the third node, which provides 'links to notification preferences and settings', but again does not address the enabling of notifications. The fourth node describes 'the alert menu and notification display', and the fifth node explains 'how to get an RSS feed of notifications and adjust notification settings', both of which also miss the crucial point of enabling notifications in the wiki settings. Thus, all nodes are ranked incorrectly as they do not provide relevant information to the input query.

ContextualRecall: The score is 0.50 because while the 1st node in the retrieval context supports the mention of 'notifications' and 'wiki' in the first sentence of the expected output, none of the nodes provide information about 'xwiki.properties' or setting 'notifications.enabled' to 'true', which is necessary for the second sentence.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications through the 'xwiki.properties' file or setting 'notifications.enabled' to 'true'.

Correctness: The actual output does not directly instruct on setting 'notifications.enabled' to 'true' in 'xwiki.properties' as the expected output does. It instead suggests checking configuration, permissions, and technical errors without mentioning the specific property setting.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.6145880069101199

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 0.875

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Correctness: 0.26252804146071973

Reasons:

AnswerRelevancy: The score is 0.75 because the response partially addresses the issue of not receiving notifications, but includes irrelevant statements such as source links, syntax documentation, and image links that do not directly help solve the problem.

Faithfulness: The score is 0.88 because the actual output incorrectly claims that notifications might be disabled for all users by default, whereas the retrieval context states that administrators have the ability to enable notifications for all users by default.

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than irrelevant nodes. The first node discusses 'Enable and disable the extension', and the second node covers 'Follow a user', both directly related to the input about enabling notifications. The irrelevant nodes, such as the third node about 'getting a notification RSS feed and settings' and the fourth node listing 'various notification preferences macros', are correctly ranked lower as they do not directly address the input needs.

ContextualRecall: The score is 0.50 because while some general concepts like enabling notifications and following users or pages (related to nodes in retrieval context) are supported, specific details such as setting 'notifications.enabled' in 'xwiki.properties', clicking the 'alert' menu, and viewing followed users in the 'network' tab are not found in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' to true, following pages, and using the alert menu.

Correctness: The actual output addresses notifications but lacks the specific steps in the expected output, such as setting 'notifications.enabled' to 'true' in 'xwiki.properties' and using the 'alerte' menu. It also introduces new elements like RSS feeds and macros not mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8540213394128188

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Correctness: 0.43841375076262723

Reasons:

AnswerRelevancy: The score is 0.79 because the response partially addresses the question about enabling notifications for one's own actions, but includes irrelevant information such as unrelated source links and discussions about a 'No Notification' tag.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which explains 'By default, you won't receive notifications for actions done by yourself. This can be changed

by switching off the Own event filter,' is ranked first, while all irrelevant nodes are ranked lower. Great job on getting it perfectly right!

ContextualRecall: The score is 1.00 because the expected output is fully supported by the nodes in the retrieval context, with no discrepancies. Great job!

CustomContextualRelevancy: The retrieval context confirms users don't receive notifications for their own actions by default and mentions the ability to change this setting, aligning closely with the expected output. However, it does not explicitly mention 'Options de filtrage avancées' for changing settings.

Correctness: Actual output provides steps for notification settings not detailed in expected output and contains additional information on managing filters, which is not present in expected output. Both outputs mention default notification settings and the ability to modify them.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7372902344420788

Individual_scores:

AnswerRelevancy: 0.5833333333333334

Faithfulness: 0.8333333333333334

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Correctness: 0.20707473998580586

Reasons:

AnswerRelevancy: The score is 0.58 because the actual output contains several irrelevant statements, such as source links and image links, which do not address the question about disabling notifications. However, the score is not lower because there may still be some relevant information present that partially addresses the input question.

Faithfulness: The score is 0.83 because the actual output incorrectly suggests that 'Create global notification filters' is used to manage global notifications, whereas the retrieval context clarifies that it is only for creating global notification filters, not managing them.

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant node. The irrelevant node, ranked fourth, is about 'getting a notification RSS feed' and does not relate to 'modifying notification settings or automatic page watching', which is less relevant compared to the first three nodes that directly address 'auto-watched mode', 'Notifications Auto Watch Preferences', and 'Auto Watch' settings.

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context covers aspects like auto-watched pages and custom filters, aligning with the expected answer about automatic page tracking and filter management. However, it lacks explicit mention of modifying 'Suivi automatique des pages' in user profile settings.

Correctness: The actual output lacks specific mention of 'following pages with significant changes' and 'deleting existing notification filters', which are key facts in the expected output. It introduces different instructions on managing notifications not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.22798064606748236

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 0.42857142857142855

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.1893124478334656

Reasons:

AnswerRelevancy: The score is 0.75 because the response addresses the main question about configuring XWiki for automatic user activation with OpenID Connect, but it includes repeated irrelevant source links that do not directly resolve the configuration issue.

Faithfulness: The score is 0.43 because the actual output incorrectly associates the setting `xwiki.authentication.createuser=true` with OpenID Connect, while the retrieval context does not specify this association. Additionally, the output mistakenly claims that new users are configured as 'context user' by default, contradicting the context which states users are created only if `xwiki.authentication.createuser` is set. Furthermore, the output incorrectly suggests setting `xwiki.authentication.authclass` for OpenID Connect user creation, whereas the context indicates this is for Container Authentication.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based authentication and cookie settings,' which is unrelated to enabling user accounts automatically with OpenID Connect. The second node covers 'container authentication and LDAP,' not mentioning the specific configuration needed for OpenID Connect user activation. The third node, while mentioning OpenID Connect, focuses on 'general authentication and configuration guides,' not on enabling user accounts automatically. The fourth node lists 'dependencies for the OpenID Connect extension,' which is not relevant to the specific configuration needed for automatic user activation. Lastly, the fifth node is about 'LDAP and Active Directory integration,' which does not relate to OpenID Connect user account activation.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention the `xwiki.properties` file or the `oidc.enableUser` property, which are crucial for the expected answer.

Correctness: Actual output focuses on account creation, not activation, and uses `xwiki.cfg` instead of `xwiki.properties`. It mentions different properties and steps than expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.523241140499706

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Correctness: 0.43944684299823616

Reasons:

AnswerRelevancy: The score is 1.00 because the response is fully relevant and directly addresses the question about customizing user profile fields from an OpenID Connect provider in XWiki. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management class and endpoints for OpenID Connect', the second node is about 'container authentication and other methods like LDAP', the third node describes 'the OpenID Connect Authenticator extension', the fourth node lists 'dependencies for the OpenID Connect Authenticator extension', and the fifth node provides 'links to configuration guides and describes limitations'. None of these contexts address user profile field customization or mapping, which is the focus of the input.

ContextualRecall: The score is 0.50 because while the retrieval context nodes mention 'xwiki.properties' and 'oidc.user.mapping', which support sentences 1 and 2 of the expected output, they do not cover specific details like 'myxproperty1', 'myxproperty2', or the steps of saving and restarting XWiki mentioned in sentences 3 and 4.

CustomContextualRelevancy: The retrieval context mentions OpenID Connect and xwiki.properties but lacks details on customizing user profile fields or using oidc.user.mapping as described in the expected output.

Correctness: Actual output provides a general method for mapping user properties but lacks specific details on using 'xwiki.properties' and 'oidc.user.mapping' as mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5705551457837964

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.4444444444444444

CustomContextualRelevancy: 0.5

Correctness: 0.28841023978214275

Reasons:

AnswerRelevancy: The score is 0.86 because the response mostly addresses the process of configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes some irrelevant details about modifying and saving user profiles, which are not directly related to the main topic of group synchronization.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which provides information about group synchronization and mentions the use of claims and the 'oidc.groups.mapping' property, is ranked third. However, irrelevant nodes discussing topics like 'authentication management and endpoints', 'container authentication', 'general OpenID Connect configuration', and 'dependencies for the OpenID Connect extension' are ranked higher, which affects the score negatively.

ContextualRecall: The score is 0.44 because while the 3rd node in the retrieval context provides a general overview of group synchronization and relevant properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', it lacks specific instructions found in the expected output, such as opening the file with a text editor, replacing placeholders with actual values, and saving and restarting XWiki.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using oidc.groups.mapping, but lacks specific details about configuring oidc.groups.claim and oidc.groups.separator properties as outlined in the expected output.

Correctness: The actual output addresses synchronization and mapping, but lacks details on configuring `oidc.groups.claim` and `oidc.groups.separator`, and omits instructions about saving and restarting XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.30702494230617144

Individual_scores:

AnswerRelevancy: 0.6428571428571429

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.1992925109798859

Reasons:

AnswerRelevancy: The score is 0.64 because the actual output contains multiple irrelevant statements in the form of source links that do not address the specific question about restricting file uploads, which lowers the score. However, the relevant parts of the answer do provide some useful information related to the question, maintaining the score above average.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'language preferences and translations in XWiki', which is unrelated to file upload restrictions. The second node describes 'various XWiki applications and their functionalities', without mentioning file upload settings. The third node is about 'CSS and XSL transformations for PDF exports', which does not relate to file upload restrictions. The fourth node lists 'various XWiki macros and their descriptions', none of which are relevant to configuring file upload restrictions. The fifth node contains 'image links and references', which do not provide information about configuring file upload settings. Similarly, the sixth node consists of 'image links and references', which are unrelated to the topic of file upload restrictions.

ContextualRecall: The score is 0.00 because none of the expected output content regarding file type or size restrictions, MIME type configuration, or attachment size settings is supported by information in the nodes in retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about configuring MIME types or setting maximum upload size as described in the expected output.

Correctness: The actual output mentions steps for image and PDF upload restrictions but lacks specific configuration details like MIME type settings and exact steps for setting max upload size as in the expected output. It also introduces additional content about plugins and customization not present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.4372880377309534

Individual_scores:

AnswerRelevancy: 0.625

Faithfulness: 0.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.2153948930523874

Reasons:

AnswerRelevancy: The score is 0.62 because the response partially addresses the question about refusing script rights to a space administrator, but includes irrelevant source links that do not directly contribute to answering the query.

Faithfulness: The score is 0.00 because the actual output incorrectly suggests changing the 'Script' permission to 'DENY' for all users, while the retrieval context already states that the 'Script' right is set to DENY by default, making the suggestion unnecessary.

ContextualPrecision: The score is 0.58 because the first node in the retrieval context is about 'language preferences and translations in XWiki', which does not relate to script rights or administrative permissions, yet it is ranked higher than relevant nodes. Additionally, the fourth node discusses 'checking access rights programmatically' and the fifth node is about 'executing code with specific user rights', neither of which directly address the issue of refusing script rights to an administrator. However, the second and third nodes are correctly identified as relevant, as they discuss 'programming and script rights' and provide a 'detailed list of rights', supporting the expected output.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the retrieval context, showing a clear understanding and accurate recall. Great job!

CustomContextualRelevancy: The retrieval context mentions that admin rights are not deniable, but it lacks specific information about XWiki not supporting the refusal of rights implied by admin rights.

Correctness: The actual output provides steps to modify script rights, which contradicts the expected output stating admin rights cannot be denied in XWiki. No mention is made of admin rights' immutability as highlighted in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.33429012576248934

Individual_scores:

AnswerRelevancy: 0.7333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.27240742124160283

Reasons:

AnswerRelevancy: The score is 0.73 because the answer addresses the configuration question about XWiki, but it includes multiple irrelevant source links that do not contribute to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'a specific page in XWiki and its metadata', which does not relate to configuring access rights for team spaces. The second node talks about 'modifying resources using XML and HTTP methods', which is unrelated to setting access permissions in XWiki. The third node is about 'managing comments and attachments on XWiki pages', which does not pertain to access control configurations. The fourth node explains 'the concept of spaces and pages in XWiki', but does not provide information on configuring access rights for team spaces. The fifth node discusses 'actions needed after running a migrator and changes in XWiki APIs', which is not relevant to setting access

permissions. Finally, the sixth node covers 'configuration display options in XWiki', which does not directly relate to setting access rights for team spaces.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, indicating a complete lack of relevant information.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting rights at the wiki or space level, or inheriting rights for child pages as described in the expected output.

Correctness: The actual output lacks specific steps like creating groups and setting rights at the wiki level. It introduces 'scope' and 'Contrôle d'accès' parameters not mentioned in the expected output, and doesn't mention inheriting rights for child pages.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7293358003717917

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Correctness: 0.6950624212783685

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the question about authenticating users with access tokens, but includes multiple irrelevant source links and section headers that do not contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because relevant nodes in the retrieval context are not consistently ranked higher than irrelevant ones. The first node, ranked 1, is irrelevant as it discusses 'LDAP configuration' and lacks mention of 'access tokens or JWT authentication.' The second node, ranked 2, is relevant because it covers 'Authorized Applications Configuration,' which pertains to using tokens for authentication. The third node, ranked 3, is irrelevant since it focuses on 'OpenID Connect scopes and response types,' not directly related to 'access tokens for authentication.' The fourth node, ranked 4, is relevant as it details 'JWT for authentication,' directly aligning with using access tokens. The fifth node, ranked 5, is irrelevant because it talks about 'different authentication mechanisms in XWiki' without specifically addressing 'access tokens or JWT.'

ContextualRecall: The score is 0.71 because while several key aspects of the expected output, such as the use of JWTs and configuration steps, are well-supported by nodes in the retrieval context, specific details like generating a JWT token with required claims and including it in the 'Authorization' header are not directly mentioned.

CustomContextualRelevancy: The retrieval context includes information about enabling the token authenticator, configuring authorized applications, and JWT token requirements, which aligns with the expected output. However, it lacks specific details about claims like expiration time, issuer, subject, and audience, and the use of the 'Authorization' header.

Correctness: The actual output correctly describes enabling the token authenticator and mentions user creation based on token information, aligning with expected output. However, it misses details on token claims, Ed25519 key signing, and configuring authorized applications.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.23166351180656078

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 0.5

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.12075030160859557

Reasons:

AnswerRelevancy: The score is 0.77 because the response partially addresses the question about authenticating users with access tokens, but includes irrelevant source links that do not directly answer the query.

Faithfulness: The score is 0.50 because the actual output incorrectly claims that the Model-Validation-UI extension is used for setting up access tokens and configuring authentication scenarios, which are not mentioned in the retrieval context. Additionally, it describes steps involving the Name Strategy Scheduler and Model-Validation API for access tokens, which are not supported by the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'model validation strategies in XWiki', which is unrelated to 'user authentication with access tokens'. The second node is about 'attachment validation and configuration in XWiki', not 'authentication using access tokens'. The third node provides information on 'the admin UI for model validation', which does not pertain to 'authenticating users with access tokens'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, as the context lacks any mention of relevant concepts such as 'Benutzer mit Zugriffstoken', 'LLM-Anwendung (BETA)-Erweiterung', 'JWT', or related authentication processes.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT as described in the expected output.

Correctness: The actual output discusses authentication with access tokens but does not match the expected JWT-based token authentication method. Key steps like configuring `xwiki.authentication.authclass` and using JWT with specific claims are missing. Instead, it mentions Model-Validation-UI, which is not part of the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.33769072364534414

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.24836656409428706

Reasons:

AnswerRelevancy: The score is 0.78 because the response mostly addresses the question about authenticating users with access tokens, but includes irrelevant source links that do not directly contribute to the explanation.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in retrieval context are irrelevant. The first node discusses 'upgrading XWiki' without mentioning 'authentication or access tokens'. The second node is about 'user management features' and does not relate to 'authentication using access tokens'. The third node focuses on 'managing user rights and profiles', with no mention of 'token-based authentication'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are key elements in the expected answer.

Correctness: The actual output discusses authentication using tokens in XWiki but lacks specific details like JWT claims, Ed25519 signing, and configuration steps mentioned in the expected output. It also includes irrelevant aspects like user profile management, which are not part of the expected criteria.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_phi3_mini-128k_4b_Q4

File: qa_001_result.json

Overall_score: 0.16666666666666666

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because none of the statements in the actual output address the specific issue of the 'Bell' not being available in XWiki. Instead, the output contains unrelated information, such as source links, incomplete sentences, and topics unrelated to the XWiki 'Bell' issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'installation instructions and prerequisites' without addressing the 'Bell' icon issue. The second node describes the 'functionality and appearance of the notifications bell' but lacks information on resolving its absence. The third node lists 'dependencies' unrelated to the 'Bell' icon problem, and the fourth node focuses on 'customizing notification templates' without mentioning the 'Bell' icon issue.

ContextualRecall: The score is 0.00 because there is no information in the nodes in retrieval context about enabling notifications in the wiki or the specific setting `notifications.enabled` in `xwiki.properties`.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the `notifications.enabled` setting in `xwiki.properties`, which is the key information in the expected output.

Correctness: The actual output does not mention 'notifications.enabled' or 'xwiki.properties', and provides irrelevant content unrelated to enabling notifications.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6648983710202968

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8888888888888888

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.6

Correctness: 0.25050133723289186

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the issue of not receiving notifications in XWiki without any irrelevant information. Great job!

Faithfulness: The score is 0.89 because the actual output incorrectly claims that user interaction is required before any activity triggers a notification in XWiki 15.5+, whereas the retrieval context does not mention this requirement.

ContextualPrecision: The score is 0.75 because the first node in the retrieval context is relevant, explaining 'how to follow a user to receive notifications', and the fourth node also provides 'an overview of different notification methods in XWiki'. However, the second node, ranked higher than the fourth, discusses 'filters and settings for notifications' without directly addressing enabling notifications, and the third node, also ranked higher than the fourth, mentions 'RSS feeds and notification settings' but lacks crucial information on enabling notifications. Additionally, the fifth node talks about the 'What's New' application, which is not directly related to the expected output.

ContextualRecall: The score is 0.50 because while there are references in the nodes in retrieval context to clicking on the 'alert' menu, following a user, and the 'network' tab in the user profile (supporting sentences 4, 5, and 6), there is no mention of enabling notifications in 'xwiki.properties', subscribing to pages or users, or watching a page, its children, or an entire wiki (missing support for sentences 1, 2, and 3).

CustomContextualRelevancy: The retrieval context includes information on following users and receiving notifications, but lacks details on enabling notifications in xwiki.properties and watching pages or wikis.

Correctness: The actual output does not mention enabling notifications in 'xwiki.properties', watching pages or users, or the 'alert' menu. It includes additional information on notification types, RSS feeds, and software updates not in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.6851851851851851

Individual_scores:

AnswerRelevancy: 0.1111111111111111

Faithfulness: 1

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.11 because the output contains numerous irrelevant statements that do not address the question about enabling notifications for one's own actions in XWiki. The presence of these unrelated details significantly detracts from the relevancy of the response.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant nodes in the retrieval context are ranked higher than the irrelevant nodes. The first node mentions that 'By default, one don't receive notification about one's own activity... You have the ability to disable this filter in the "Advanced filtering options" section of your own notification settings,' which directly relates to the expected output. The second node states 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' aligning with the expected output about disabling the filter. The irrelevant nodes, such as the third node that focuses on administrators configuring notifications and does not mention user-specific settings, are ranked lower. Great job on the perfect ranking!

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the information in the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context fully contains the expected answer's information about default notification settings and the ability to change them in the advanced filtering options of the user profile.

Correctness: The actual output contains no relevant information about notifications settings or disabling filters as mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.3380384813101436

Individual_scores:

AnswerRelevancy: 0.2222222222222222

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Correctness: 0.006008665638639484

Reasons:

AnswerRelevancy: The score is 0.22 because the actual output contains multiple irrelevant statements that do not address the question about disabling notifications in XWiki, such as discussing unrelated topics like Python code snippets and SEO. However, the score is not zero, indicating there might be some relevant information present.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'getting a notification RSS feed and the settings for notifications', but it does not mention disabling notifications or automatic page watching. The second node explains 'how to follow a user and default watch settings', but lacks information on disabling notifications or changing automatic page watching settings. Subsequent nodes similarly fail to address the user's query about disabling notifications or automatic page watching, focusing instead on unrelated aspects like 'filtering notifications', 'administrator level settings', and 'overriding default notification templates'.

ContextualRecall: The score is 0.50 because while the retrieval context supports the concept of automatic page watching and changing notification settings in the user profile (as seen in nodes 1 and 2 in the retrieval context), it lacks information about being notified of changes on your work and deleting

existing notification filters, which are also part of the expected output.

CustomContextualRelevancy: The retrieval context mentions default watch settings and user profile settings but lacks details on automatic page watching for major modifications and deleting notification filters.

Correctness: The actual output contains irrelevant information and lacks key facts present in the expected output such as automatic page watching and notification settings.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.6019607843137255

Individual_scores:

AnswerRelevancy: 0.011764705882352941

Faithfulness: 1

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.01 because the output contains numerous irrelevant statements focusing on LDAP authentication and documentation, which do not address the specific OpenID Connect configuration issue in XWiki. The minimal relevance is due to the presence of some authentication-related content, but it lacks direct applicability to the question asked.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions '# oidc.enableUser=true', is correctly ranked first. This node is directly relevant to enabling user accounts on first login, aligning perfectly with the input question. Great job on the ranking!

ContextualRecall: The score is 0.60 because while the node(s) in retrieval context correctly identify the property '# oidc.enableUser=true' that needs to be adjusted in the xwiki.properties file (as seen in sentences 1, 2, and 3 of the expected output), they do not mention the necessity of saving the file and restarting the XWiki instance, nor do they explicitly state that new user accounts will be automatically enabled upon first login via OpenID Connect, which are covered in sentences 4 and 5.

CustomContextualRelevancy: The retrieval context includes the exact property '# oidc.enableUser=true' mentioned in the expected output, indicating that the necessary information to enable user accounts on first login is present.

Correctness: The actual output focuses on LDAP and Azure Active Directory configurations, which do not relate to the expected output about enabling user accounts via OpenID Connect in XWiki by modifying the 'xwiki.properties' file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.6666666666666666

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because none of the statements in the actual output address the question about customizing user profile fields in XWiki. Instead, they discuss unrelated topics such as OAuth authentication in Django, SQLAlchemy models, and other irrelevant subjects.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user', is ranked first. This directly addresses the input question about customizing user profile fields. The irrelevant nodes, discussing topics like 'access tokens and scopes', 'endpoints and HTTP methods', 'group synchronization and customization templates', and 'in-wiki configuration and multiple configurations through cookies', are all ranked lower, ensuring a perfect contextual precision score. Great job on getting it spot on!

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the information in the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context includes detailed information on customizing user profile fields using the `oidc.user.mapping` property in the `xwiki.properties` file, matching the expected answer.

Correctness: The actual output does not address any elements of the expected output; it discusses OAuth and Django instead of OpenID Connect and XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8413490470217143

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 1.0

Correctness: 0.5480942821302864

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider, with no irrelevant information included. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant nodes. The first three nodes provide detailed information on configuring properties like `oidc.groups.claim`, `oidc.groups.mapping`, and `oidc.groups.separator`, which are crucial for setting up group synchronization. The irrelevant nodes, ranked fourth and fifth, focus on prerequisites, installation instructions, and general authentication management, which do not directly address the group synchronization setup process. Great job on achieving perfect precision!

ContextualRecall: The score is 0.50 because while specific configuration details like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' are supported by nodes in the retrieval context, the general instructions for locating, editing, and saving the 'xwiki.properties' file are not directly found in

the retrieval context.

CustomContextualRelevancy: The retrieval context contains all key information from the expected output, including configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' properties for group synchronization in the 'xwiki.properties' file.

Correctness: The actual output includes additional steps like enabling group synchronization, custom group synchronization, and troubleshooting that are not in the expected output. It lacks the specific details about configuring the separator and saving the file as in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6082865190605594

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 0.8333333333333334

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.8181818181818182

CustomContextualRelevancy: 0.3

Correctness: 0.44820396284820496

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is highly relevant to the question about configuring attachment size limits, but it includes an unnecessary link to a general guide on XWiki basic concepts, which slightly detracts from its directness.

Faithfulness: The score is 0.83 because the actual output incorrectly specifies that the 'Maximum Upload Size' field default is set in bytes for XWiki versions before 10.9RC1, while the retrieval context only mentions a default of about 32 megabytes without linking it to any specific version.

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which provides information about setting limits on the maximum size of an attachment and mentions the XWikiPreferences document, is ranked third. However, the first and second nodes discuss 'linking to external files and backlinks indexation' and 'the structure of XWiki pages and available actions', which are unrelated to configuring upload restrictions or size limits, and should be ranked lower than the relevant node.

ContextualRecall: The score is 0.82 because most of the expected output sentences align well with the nodes in the retrieval context, particularly regarding configuring mimetypes and setting maximum attachment size. However, the introductory sentence and specific configuration line are not directly supported by any node in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific details like the 10MB limit and exact mimetype configuration steps found in the expected output.

Correctness: Actual output mentions configuring attachment size to 10MB correctly but uses 'xwiki.cfg' instead of 'xwiki.properties' for mimetype configuration and omits the mimetype restriction step.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.5007014621588045

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.004208772952827291

Reasons:

AnswerRelevancy: The score is 0.00 because the output contains numerous irrelevant statements that do not address the question about denying script rights to a space administrator.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses programming rights and admin rights implications, but it does not address denying script rights to a space administrator. The second node lists various rights and their default states without mentioning the specific denial of script rights. The third node explains script rights and their default status but lacks information on denying them. The fourth node provides technical details about checking access rights but not about denying script rights. Lastly, the fifth node discusses converting rights and excluding pages or spaces, which is unrelated to denying script rights to a space administrator.

ContextualRecall: The score is 1.00 because the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context confirms that rights implied by admin cannot be denied, matching the expected output.

Correctness: The actual output is entirely unrelated to the expected output, containing irrelevant and incoherent content that does not address the topic of XWiki rights.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.6484931667977797

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 0.7777777777777778

ContextualPrecision: 0.6666666666666666

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Correctness: 0.28937169919937694

Reasons:

AnswerRelevancy: The score is 0.86 because the response largely addresses the question about configuring permissions in XWiki, but includes sources that do not directly relate to the specific task of configuring permissions, which slightly detracts from its relevance.

Faithfulness: The score is 0.78 because the actual output incorrectly claims that actions can be allowed/denied at the wiki-wide level and page level by selecting 'Space', which contradicts the retrieval context that does not specify selecting 'Space' for these levels.

ContextualPrecision: The score is 0.67 because relevant nodes in the retrieval context, such as the first node discussing 'rights inheritance and administration features' and the fourth node explaining 'the permission system in XWiki', are ranked higher. However, irrelevant nodes like the second node focusing on 'parent/child relationship and the Nested Pages concept' and the third node about 'string formatting and user identifiers in OpenID Connect' are ranked higher than some relevant nodes, which prevents a higher score.

ContextualRecall: The score is 1.00 because every sentence in the expected output is well-supported by the nodes in the retrieval context. Fantastic job ensuring all details are covered!

CustomContextualRelevancy: The retrieval context mentions rights inheritance and setting rights at different levels, but lacks specific steps like creating groups for each team or setting space-level rights as detailed in the expected output.

Correctness: The actual output incorrectly suggests setting both view and edit permissions to Deny at the page level, conflicting with the expected output's guidance of allowing view access to all and restricting edit access to each team's space. It also introduces unnecessary steps not present in the expected output, like setting deny permissions for creating pages outside of assigned spaces.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.32449659982205725

Individual_scores:

AnswerRelevancy: 0.27906976744186046

Faithfulness: 1

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.0012431648238162745

Reasons:

AnswerRelevancy: The score is 0.28 because the majority of the statements in the output are unrelated to the issue of the missing notification bell icon, focusing instead on various unrelated topics such as guides, Python functions, and notification settings in different contexts. This lack of relevance significantly lowers the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node in the retrieval context, which is the sixth node, 'mentions editing the xwiki.properties file, which is relevant to enabling notifications by setting `notifications.enabled` to `true`.' However, it is ranked lower than five irrelevant nodes. For example, the first node 'discusses custom notification filters and their behavior, which does not relate to enabling notifications in the wiki settings,' and should be ranked lower than the relevant node.

ContextualRecall: The score is 0.50 because while the expected output partially aligns with the retrieval context, specifically regarding editing the xwiki.properties file (node 6 in retrieval context), it lacks direct support for the specific mention of 'notifications.enabled' and setting it to 'true'.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via `xwiki.properties` or setting `notifications.enabled` to `true`.

Correctness: The actual output does not contain any relevant information about enabling notifications in the wiki settings as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_012_result.json

Overall_score: 0.16666666666666666

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because none of the statements in the actual output are relevant to the issue of not receiving notifications. All statements address unrelated topics, such as guides on various subjects and AI technology, which do not pertain to the input query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node contains 'links and images related to various references, which do not provide any information about notifications or how to enable them.' The second node is also irrelevant as it 'consists of links and images related to different references, which are not relevant to the topic of notifications or their settings.' Similarly, the third node includes 'links and images that are unrelated to the issue of receiving notifications or adjusting notification settings.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, which primarily contains image links and references.

CustomContextualRelevancy: The retrieval context contains image links and references unrelated to the notification settings and following features described in the expected output.

Correctness: The actual output is completely unrelated to the expected output content about enabling notifications and following users or pages in XWiki. It instead discusses various topics on AI and healthcare, none of which align with any aspect of the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_013_result.json

Overall_score: 0.16666666666666666

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because none of the statements in the actual output address the issue of enabling notifications for one's own actions. All statements are irrelevant to the input question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customization of page templates, internationalization, and other features unrelated to notification settings.' The second node is about 'administrating wiki features like users, groups, and rights, which is not relevant to notification settings for user activities.' The third node covers 'extensions like Office Importer, Font Awesome Icon Theme, and others, none of which relate to user notification settings.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be linked to any nodes in the retrieval context, as there is no information about notifications, triggering events, or advanced notification settings in the nodes.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters as mentioned in the expected output.
Correctness: The actual output contains irrelevant and unrelated information compared to the expected output, which is about notification settings. There is no alignment in terms of content or factual statements between the actual and expected outputs.
Question_language: de
Expected_answer_language: de
Answer_language: en

File: qa_014_result.json

Overall_score: 0.2991779408587254
Individual_scores:
AnswerRelevancy: 0.7727272727272727
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.02234037242507978
Reasons:

AnswerRelevancy: The score is 0.77 because the output partially addresses the question about disabling notifications, but includes several irrelevant links to documentation pages that do not directly help with the issue.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files' which is unrelated to notification settings. The second node talks about 'moving attachments', which also does not relate to disabling notifications. Similarly, the third node lists 'various macros', none of which pertain to notification settings. The fourth node is about 'administrating wiki features', which does not mention notification settings. The fifth node discusses 'controlling permissions', again unrelated to notifications. Lastly, the sixth node lists 'XWiki extensions', none of which are related to notification settings. All these irrelevant nodes are ranked higher than any relevant content, resulting in a score of 0.00.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to automatic page following or notification settings as mentioned in the expected output.

Correctness: The actual output does not match any key facts from the expected output and includes unrelated instructions about deactivating notifications in XWiki.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_015_result.json

Overall_score: 0.3490016481967781
Individual_scores:
AnswerRelevancy: 0.7777777777777778
Faithfulness: 0.875
ContextualPrecision: 0
ContextualRecall: 0.4
CustomContextualRelevancy: 0.0

Correctness: 0.04123211140289102

Reasons:

AnswerRelevancy: The score is 0.78 because the response partially addresses the user's question about configuring automatic activation of user accounts with OpenID Connect, but includes irrelevant details about limitations and bypassing the authenticator, which do not directly help in solving the issue.

Faithfulness: The score is 0.88 because the actual output incorrectly suggests setting the session cookie on a top-level domain for all wikis, whereas the retrieval context clarifies that this approach is ineffective if the wikis have very different domains.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to configuring automatic user activation on first login. The first node discusses 'domain-based authentication and session cookies', which is unrelated. The second node provides 'general information about OpenID Connect authentication', without addressing automatic user activation. The third node describes 'OpenID Connect authenticator and its compatibility', but does not mention user activation settings. The fourth node lists 'various OpenID Connect configuration properties', yet omits the specific property `oidc.enableUser` for automatic user activation.

ContextualRecall: The score is 0.40 because while the 4th node in the retrieval context supports the mention of `xwiki.properties` and `oidc.enableUser`, there is a lack of direct information about adding or updating the property, saving the file, restarting XWiki, and ensuring automatic activation of new user accounts upon first login.

CustomContextualRelevancy: The retrieval context lacks any mention of adjusting the `xwiki.properties` file or the specific property `oidc.enableUser` to automatically activate user accounts upon first login with OpenID Connect.

Correctness: The actual output does not mention the required changes in `xwiki.properties` or the `oidc.enableUser` property, and instead focuses on unrelated configurations and troubleshooting steps, which do not align with the expected output instructions.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.4303996894416848

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.3073981366501085

Reasons:

AnswerRelevancy: The score is 0.88 because the response is mostly relevant and addresses the question about customizing user profile fields in XWiki using OpenID Connect. However, it includes source links that do not directly contribute to answering the question, which slightly lowers the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating the actual output perfectly aligns with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect provider endpoints', but does not mention 'customizing user profile fields or mapping them'. The second node talks about 'different authentication methods like Container Authentication and Active Directory', but does not cover 'customizing or mapping OpenID Connect user profile fields'. The third node provides 'links and descriptions related to OpenID Connect authentication', but lacks information on 'mapping user profile fields or using `oidc.user.mapping`'. The fourth node describes the 'OpenID Connect Authenticator

extension and its compatibility', but does not address 'customization of user profile fields or mapping them'. The fifth node lists 'dependencies for the OpenID Connect Authenticator extension', which is unrelated to 'customization of user profile fields or the use of `oidc.user.mapping`'.

ContextualRecall: The score is 0.40 because while the nodes in the retrieval context mention 'xwiki.properties' and restarting XWiki, they do not cover specific details about 'oidc.user.mapping', mapping examples, or replacing variables, which are crucial for the expected output.

CustomContextualRelevancy: The retrieval context lacks information about customizing user profile fields using `oidc.user.mapping` in `xwiki.properties`, which is detailed in the expected output.

Correctness: The actual output discusses configuring OpenID Connect but lacks specific details on mapping properties in `xwiki.properties` and restarting XWiki, which are critical in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.3700853497226966

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.22051209833617924

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about setting up group synchronization between XWiki and an OpenID Connect provider, with no irrelevant information included. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect endpoints' but lacks information on 'group synchronization or the specific configuration steps required for it.' The second node covers 'container authentication and authentication extensions' without mentioning 'group synchronization or relevant configuration details.' The third node provides a 'general description of OpenID Connect authentication and mentions automatic synchronization,' yet it does not include 'specific instructions or details about configuring group synchronization.' The fourth node lists 'dependencies for the OpenID Connect extension' but omits 'information related to group synchronization or its configuration.' Finally, the fifth node gives 'general information about the OpenID Connect Authenticator,' focusing on 'compatibility and installation details' without addressing 'group synchronization specifics.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context. The retrieval context lacks any mention of setting up group synchronization between XWiki and OpenID Connect providers, as well as all specific configuration instructions and examples provided in the expected output.

CustomContextualRelevancy: The retrieval context lacks details on configuring group synchronization properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' as specified in the expected output.

Correctness: The actual output includes configuration steps for OpenID Connect, but lacks specific details like configuring `xwiki.properties`, setting `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` as outlined in the expected output. There is no mention of saving changes and restarting XWiki.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_018_result.json

Overall_score: 0.40453861081194825
Individual_scores:
AnswerRelevancy: 0.02564102564102564
Faithfulness: 1
ContextualPrecision: 0.2
ContextualRecall: 0.7
CustomContextualRelevancy: 0.5
Correctness: 0.0015906392306643102
Reasons:

AnswerRelevancy: The score is 0.03 because the output contains numerous irrelevant statements that do not address the question about uploading specific file types with size restrictions. The statements cover unrelated topics such as AI, documentation, and guides, which significantly detracts from the relevancy of the response.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.20 because the relevant node in the retrieval context, which provides 'detailed instructions on setting the maximum size of an attachment in XWikiPreferences', is ranked fifth. Irrelevant nodes, such as the first node discussing 'trusted domains and PDF export customization', the second node about 'ODT and PDF export features', and the third node listing 'various macros and features', should be ranked lower as they do not address configuring attachment types and size limits.

ContextualRecall: The score is 0.70 because while the 5th node in the retrieval context provides detailed steps for configuring the maximum attachment size, aligning well with sentences related to this in the expected output, there is no mention of configuring allowed MIME types or editing the xwiki.properties file, which are key parts of the expected output.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and MIME type restrictions, but lacks specific details on configuring xwiki.properties or setting size to 10 MB.

Correctness: The actual output is completely unrelated to the expected output, discussing topics such as Python scripts and AI models instead of configuring MIME types and attachment sizes.

Question_language: de
Expected_answer_language: de
Answer_language: en

File: qa_019_result.json

Overall_score: 0.34622268820496704
Individual_scores:
AnswerRelevancy: 0.9259259259259259
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Correctness: 0.1514102033038764
Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and effectively addresses the input question about denying script rights to a domain administrator. However, the inclusion of a source link, which is not directly relevant, slightly detracts from a perfect score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'form validation and JavaScript patterns', which does not relate to denying script rights in XWiki. The second node covers 'template providers and page creation', which is unrelated to script rights or administrative permissions. The third node mentions 'a custom right check and configuration', but it does not address the specific issue of denying script rights to a space administrator.

ContextualRecall: The score is 0.00 because the expected output sentence about the limitation of admin rights in XWiki is not supported by any information in the nodes in retrieval context.

CustomContextualRelevancy: The retrieval context does not mention anything about admin rights or their implications, which is the main focus of the expected output.

Correctness: The actual output provides steps for configuring permissions in XWiki, which is unrelated to the expected output. The expected output states a limitation in XWiki about admin rights, which is not addressed in the actual output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.36261663365496116

Individual_scores:

AnswerRelevancy: 0.9016393442622951

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.125

CustomContextualRelevancy: 0.0

Correctness: 0.1490604576674719

Reasons:

AnswerRelevancy: The score is 0.90 because the response largely addresses the question about configuring team access in XWiki, but includes several irrelevant source links and mentions that do not directly contribute to solving the specific access configuration problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to configuring team-specific access control in XWiki. The first node discusses XWiki's database support and programming features, which are not relevant. The second node is about the Notifications Application, unrelated to access control. The third node explains Nested Pages and subwikis, not addressing team-specific access rights. The fourth node mentions managing users and groups but lacks specific information on setting team access controls. The fifth node discusses group synchronization and customization, not directly related to the input question.

ContextualRecall: The score is 0.12 because only the concept of creating groups for each team (sentence 1) is supported by node 4 in the retrieval context. The rest of the expected output sentences, such as setting wiki-wide rights, space-level rights, and inheriting rights for child pages, are not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks information about creating groups, setting rights, and inheriting rights, which are key steps in the expected output for access control in XWiki.

Correctness: Actual output suggests creating separate wikis for each team, contradicting expected output's use of groups within one wiki. It also lacks instructions for setting 'View' rights for XWikiAllGroup, which is vital in expected output to ensure visibility across teams.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.33870128996487386

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2140259216074248

Reasons:

AnswerRelevancy: The score is 0.82 because the response is mostly relevant to the input, addressing the issue with the notification feature. However, it includes links that are not directly related to resolving the specific problem, which slightly detracts from its relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'custom notification filters and their behavior' but does not address enabling notifications in the wiki settings. The second node provides 'links to various notification settings and customizations' but does not mention enabling notifications through `xwiki.properties`. The third node lists 'various notification preferences and macros' but does not provide information on enabling notifications in the wiki. The fourth node describes 'the notifications menu and its features' but does not relate to enabling notifications in the wiki configuration. The fifth node explains 'RSS feeds and settings for notifications' but does not mention enabling notifications through the `xwiki.properties` file.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about enabling notifications via the 'xwiki.properties' file or setting 'notifications.enabled' to 'true'.

Correctness: The actual output addresses issues with XWiki notifications but does not mention setting `notifications.enabled` to `true` in `xwiki.properties` as in the expected output. It introduces additional troubleshooting steps not required by the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.4389048629430968

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 0.7272727272727273

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Correctness: 0.24901359324299657

Reasons:

AnswerRelevancy: The score is 0.86 because the response is mostly relevant to the issue of not receiving notifications, but includes unnecessary links that do not directly address the problem.

Faithfulness: The score is 0.73 because the actual output inaccurately describes a 'Notifications Application - Description' page and an option to 'Enable and disable the extension', which are not mentioned in the retrieval context. Additionally, it incorrectly suggests that individual users can have notifications enabled by default, whereas the context specifies this is done for all users by administrators. Lastly, the actual output mentions enabling encryption for notification emails, which is not covered in the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node 'provides links to various sections related to notifications but does not specifically address enabling notifications or following users', which are key parts of the expected output. The second node 'lists different notification features and settings but does not directly mention enabling notifications or following users', which are essential to the expected output. The third node 'discusses RSS feeds and notification settings but does not provide information on enabling notifications or following users', which are crucial to the expected output. The fourth node 'lists different notification preferences macros but does not specifically address enabling notifications or following users', which are central to the expected output.

ContextualRecall: The score is 0.50 because while nodes in the retrieval context mention enabling extensions and following users or pages, there is no specific mention of 'xwiki.properties', the alert menu, or the 'réseau' tab, which are crucial details in the expected output.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' to 'true' and specifics on following pages or using the alert menu.

Correctness: The actual output provides detailed steps on managing notifications, including installation and customization through the Notifications Application, which is not mentioned in the expected output. However, it lacks specific instructions on following pages or users using the 'alerte' menu and does not address checking the 'notifications.enabled' property in 'xwiki.properties'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8382768146166383

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Correctness: 0.25466088769982914

Reasons:

AnswerRelevancy: The score is 0.88 because the answer mostly addresses the question about enabling notifications for one's own actions, but it includes irrelevant source links that do not directly contribute to solving the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which states 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' is ranked first. This directly addresses the input question about

receiving notifications for one's own actions, while all irrelevant nodes are ranked lower.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context confirms users don't receive notifications for their own actions by default and mentions the ability to change this setting, aligning closely with the expected output. However, it lacks explicit mention of 'Options de filtrage avancées' for disabling the filter.

Correctness: The actual output provides extensive details on enabling personal notifications, which deviates from the expected output focus on default notification settings and advanced filtering options. Key facts about default settings and advanced filtering options are missing, and additional information about macros and technical support does not align with expected content.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.5916666666666667

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because none of the statements in the output are relevant to the question about disabling notifications. Each statement addresses unrelated topics, such as URLs, Python scripts, and documentation, which do not provide the requested information.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant node. The irrelevant node, ranked fourth, discusses 'getting a notification RSS feed and the general settings menu', which does not directly address disabling automatic notifications or modifying auto-watch settings, unlike the first three nodes that focus on managing notifications and auto-watched pages.

ContextualRecall: The score is 0.75 because most sentences in the expected output can be linked to the 2nd node in the retrieval context, which discusses 'auto watched pages', 'notification preferences', and 'custom notification filters'. However, the specific mention of being notified of changes is not directly supported by the retrieval context.

CustomContextualRelevancy: The retrieval context includes information about auto-watched pages and custom notification filters, matching the expected answer. However, it lacks explicit mention of being notified of changes on one's work and modifying notification settings in the user profile.

Correctness: The actual output contains unrelated information about Python and AI, while the expected output discusses notification settings and management in a user profile context.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_025_result.json

Overall_score: 0.24146341463414633

Individual_scores:

AnswerRelevancy: 0.04878048780487805

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.05 because the majority of the statements in the actual output are irrelevant to the input question about configuring OpenID Connect in XWiki. The output primarily discusses unrelated topics such as LDAP authentication, e-commerce, AI language modeling, and Azure Active Directory, which do not address the user's specific query about automatic user account activation in XWiki using OpenID Connect.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job maintaining perfect alignment!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input query about configuring automatic user activation with OpenID Connect in XWiki. The first node discusses 'domain-based authentication and session cookies', which is not related to the query. The second node focuses on 'container authentication and LDAP', which does not address the specific configuration needed. The third node, although mentioning OpenID Connect, lacks details on 'enabling automatic user activation'. The fourth node lists 'dependencies for the OpenID Connect extension' without addressing activation settings. Lastly, the fifth node is about 'LDAP and Active Directory integration', unrelated to the desired OpenID Connect configuration.

ContextualRecall: The score is 0.40 because while nodes in the retrieval context mention 'OpenId Connect Authentication' and 'xwiki.properties file', which supports parts of sentences 1 and 2, there is no specific mention of adding or updating properties, saving the file, restarting XWiki, or automatic activation of user accounts on first login, which are key aspects of sentences 3, 4, and 5.

CustomContextualRelevancy: The retrieval context lacks information about adjusting xwiki.properties for OpenID Connect to automatically activate user accounts, specifically the property oidc.enableUser.

Correctness: The actual output is unrelated to the expected output, discussing LDAP and Azure Active Directory instead of OpenID Connect and XWiki configuration.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_026_result.json

Overall_score: 0.26666666666666666

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.2

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because none of the statements in the actual output are relevant to the question about customizing user profile fields in XWiki. The output contains numerous unrelated topics, such as setting up authentication systems and writing documents, which do not address the input query.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses authentication management and OpenID Connect endpoints but lacks information on user profile field mapping or customization. The second node focuses on container authentication and other methods like LDAP and Active Directory, which are unrelated to the input query. The third node provides general information about the OpenID Connect Authenticator extension but does not address the specific need for user profile field mapping or customization. The fourth node lists dependencies for the OpenID Connect Authenticator extension, which is not relevant to the input. Finally, the fifth node offers links to configuration guides but fails to mention mapping or customizing user profile fields in `xwiki.properties`.

ContextualRecall: The score is 0.40 because while the 1st sentence of the expected output is supported by the 1st node in the retrieval context, the rest of the expected output, including the example mapping, instructions to replace variables, and the step to save and restart XWiki, are not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions OpenID Connect and xwiki.properties but lacks details on customizing user profile fields and mapping as described in the expected output.

Correctness: The actual output does not contain any relevant information related to the expected output. There is no mention of user profile field customization, OpenID Connect, or mapping properties in `xwiki.properties`. The content is unrelated and does not maintain the context or factual details of the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_027_result.json

Overall_score: 0.5400776853645012

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.3

Correctness: 0.27379944552034013

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, ranked third, directly addresses group synchronization by mentioning the claim 'xwiki_groups' and the 'oidc.groups.mapping' property, which are crucial for configuring group synchronization. However, it is ranked lower than two irrelevant nodes. The first node focuses on authentication management and endpoints configuration, and the second node discusses different authentication methods and extensions, neither of which address group synchronization with OpenID Connect. This misordering affects the score, preventing it from being higher.

ContextualRecall: The score is 0.33 because while the 3rd node in the retrieval context supports the configuration of 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' as mentioned in sentences 3, 5, and 7 of the expected output, it lacks information on locating the 'xwiki.properties' file, providing specific examples, and saving changes as described in sentences 1, 2, 4, 6, and 8.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using 'oidc.groups.mapping', but lacks details on 'oidc.groups.claim' and 'oidc.groups.separator' properties or saving changes in 'xwiki.properties'.

Correctness: Actual output introduces concepts like `oidc.userinfoclaims` and custom group handlers not in expected output, lacks specific steps like setting `oidc.groups.claim` and `oidc.groups.separator`, but mentions `oidc.groups.mapping` and testing configuration as expected.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.32158982320128277

Individual_scores:

AnswerRelevancy: 0.7619047619047619

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.16763417730293478

Reasons:

AnswerRelevancy: The score is 0.76 because the output partially addresses the question by providing relevant information about XWiki documentation, but it is not higher due to repeated irrelevant statements that do not directly address the specific requirement of restricting file uploads to images or PDFs of a certain size.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'language preferences and translations in XWiki', which is unrelated to configuring file upload restrictions. The second node covers 'various applications and features of XWiki, such as tour guides and mail configuration', which do not relate to file upload settings. The third node is about 'PDF export customization using XSL and CSS', which is unrelated to setting file upload restrictions in XWiki. The fourth node discusses 'system filters, office document viewing, and other unrelated features', not about file upload restrictions. The fifth and sixth nodes contain 'references to various projects and images', which are not relevant to the topic of file upload restrictions.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context, as the context lacks information on file upload restrictions, MIME type configuration, and attachment size settings.

CustomContextualRelevancy: The retrieval context lacks any information related to configuring MIME types or setting attachment size limits as described in the expected output.

Correctness: The actual output does not align with the expected output. It incorrectly suggests using an extension for setting file size limits, which is not mentioned in the expected steps. It also does not specify modifying xwiki.properties for MIME types or setting the size in XWikiPreferences.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.46906995155191383

Individual_scores:

AnswerRelevancy: 0.027777777777777776

Faithfulness: 1

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Correctness: 0.0033085982003720816

Reasons:

AnswerRelevancy: The score is 0.03 because the output contains numerous irrelevant statements that do not address the question about refusing script rights to a space administrator. The low score reflects the lack of relevant content.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.58 because the first node in the retrieval context is irrelevant as it 'discusses language preferences and translations in XWiki,' which is unrelated to script rights or administrative permissions. Additionally, the fourth and fifth nodes are also irrelevant since they 'discuss technical methods for checking access rights programmatically' and 'provide information on executing tasks with specific user rights,' neither of which directly address the ability to refuse script rights to an administrator. However, the second and third nodes are relevant as they discuss 'programming and script rights' and provide 'detailed information on rights,' supporting the expected output.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context mentions that admin rights imply other rights and are not deniable, but it does not explicitly state that admin rights cannot be refused, as mentioned in the expected output.

Correctness: The actual output contains unrelated and irrelevant information compared to the expected output, which focuses on access rights in XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_030_result.json

Overall_score: 0.3461268303287232

Individual_scores:

AnswerRelevancy: 0.7333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.34342764863900593

Reasons:

AnswerRelevancy: The score is 0.73 because the response partially addresses the question about configuring spaces in XWiki, but the inclusion of multiple source links that do not directly contribute to the solution reduces its overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node provides 'metadata about a page in XWiki, but it does not address access control or permissions settings,' which is not pertinent to configuring team space access. The second node discusses 'modifying resources using XML and HTTP methods,' unrelated to the topic of access control in XWiki. The third node is about 'managing comments and attachments,' which does not relate to setting access permissions for spaces. The fourth node explains 'the creation of spaces and sub-pages in XWiki,' but lacks information on access permissions. The fifth node discusses 'migration actions and changes in XWiki,' not relevant to configuring access control for team spaces. Lastly, the sixth node describes 'configuration display settings in XWiki,' which does not address access control or

permissions for spaces.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks information about creating groups, setting rights at the wiki and space level, and inheriting rights for child pages as outlined in the expected output.

Correctness: Actual output lacks specific actions like creating groups, setting rights at wiki and space levels, and inheriting rights for child pages, which are present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.5023809523809524

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because none of the statements in the actual output are relevant to the question about authenticating users with access tokens. The output contains numerous unrelated topics and lacks any direct answer to the input question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 0.50 because relevant nodes in the retrieval context, such as the second node which 'explains the configuration of authorized applications, which is relevant to using access tokens for authentication', and the fourth node which 'provides a description of the extension for authenticating users with JWT tokens, directly addressing the input question', are ranked lower than irrelevant nodes. The first node, discussing 'LDAP configuration and does not mention token-based authentication or JWT', and the third node, focusing on 'OpenID Connect scopes and response types, which is unrelated to the use of JWT tokens for authentication', should be ranked lower than these relevant nodes.

ContextualRecall: The score is 0.71 because while several key aspects of the expected output, such as the use of JWT for authentication (sentence 1) and configuration details (sentences 2 and 5), are supported by nodes in the retrieval context, some specific instructions like generating a JWT token with required claims (sentence 3) and including it in the 'Authorization' header (sentence 4) are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context confirms the use of JWT for authentication and mentions configuring authorized applications, but lacks details on token claims and specific configuration steps like setting the authclass property.

Correctness: The actual output is completely unrelated to the expected output, discussing unrelated topics like AI and Python, whereas the expected output focuses on token-based authentication for LLM Application.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.3415142725063121

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.21575230170453938

Reasons:

AnswerRelevancy: The score is 0.83 because the main content effectively addresses the question about authenticating users with access tokens, but it includes irrelevant source links that do not contribute to answering the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'model validation strategies in XWiki', which is unrelated to 'user authentication with access tokens'. The second node is about 'attachment validation and configuration in XWiki', not about 'token-based user authentication'. The third node provides information about 'the admin UI for model validation', which does not pertain to 'authenticating users with access tokens'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are central to the expected output.

Correctness: The actual output lacks specific details like the use of Ed25519 key, and the configuration steps for 'xwiki.authentication.authclass' and 'Authorized Applications'. It also misses the note about supporting user creation or update based on token information. It provides general token-based authentication information but does not align with the expected output's specifics for LLM-Anwendung (BETA).

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.16666666666666666

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because none of the statements in the output are relevant to the question about authenticating users with access tokens. The output contains numerous irrelevant topics and statements that do not address the input question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'upgrading XWiki' and lacks any mention of 'authenticating users with access tokens'. The second node talks about 'features of the Administration Application, such as users and rights', but does not provide information on 'authenticating users with access tokens'. The third

node addresses 'managing user rights and profiles', but it does not cover 'user authentication with access tokens'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are central to the expected output.

Correctness: The actual output is unrelated to the expected output and fails to address any of the key facts or critical information outlined in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

Model: AI.Models.qa_qwen2_7b_Q4

File: qa_001_result.json

Overall_score: 0.3229596164826022

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.014680775818690081

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and addresses the issue of the 'Bell' not being available in XWiki. However, it includes a source link that does not directly address the specific problem, which slightly reduces the score.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'installation instructions and prerequisites for extensions,' which does not help with enabling notifications. The second node describes 'the notification bell and its functionality,' but lacks information on enabling it. The third node lists 'dependencies for the notifications extension,' which is not related to the issue of enabling notifications. The fourth node explains 'how to override notification templates,' without addressing enabling notifications in the wiki settings.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be linked to any nodes in the retrieval context, as there is no mention of enabling notifications or the specific settings in `xwiki.properties`.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications or the `notifications.enabled` setting in `xwiki.properties`.

Correctness: The actual output does not address enabling notifications through the `notifications.enabled` setting in `xwiki.properties` as mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6842243413933563

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.8055555555555555

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Correctness: 0.2997904928045828

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating that the actual output is perfectly aligned with the retrieval context. Great job maintaining accuracy!

ContextualPrecision: The score is 0.81 because the relevant nodes in the retrieval context are mostly ranked higher than the irrelevant ones. For example, the first node is relevant as it explains how to follow a user to receive notifications, which is pertinent to the query. However, the second node, which discusses filtering notifications and hidden pages, is not directly related to enabling notifications and should be ranked lower. Similarly, the fifth node about the 'What's New' application does not address the issue of not receiving notifications and should also be ranked lower than the relevant nodes.

ContextualRecall: The score is 0.50 because while some sentences, like those about using the 'alert' menu (node 4) and following users (node 1), are supported by nodes in the retrieval context, key details such as enabling notifications in `xwiki.properties` and subscribing to pages or users are not covered.

CustomContextualRelevancy: The retrieval context includes information on following users and receiving notifications, but lacks details on enabling notifications in xwiki.properties and watching pages or wikis.

Correctness: The actual output provides troubleshooting steps but misses specific instructions like setting `notifications.enabled` to `true` in `xwiki.properties`. It also diverges into other topics like RSS feeds, which are not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9564241922299951

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.9385451533799708

Reasons:

AnswerRelevancy: The score is 0.80 because the output is mostly relevant to the question about enabling notifications for one's own actions in XWiki, but it includes irrelevant source links that do not directly address the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant nodes in the retrieval context are perfectly ranked higher than the irrelevant ones. The first and second nodes directly address the expected output by explaining how to enable notifications for one's own actions in XWiki. In contrast, the third node focuses on administrators configuring notifications, the fourth node discusses RSS feeds and notification types, and the fifth node lists various notification methods, none of which address the specific issue of receiving notifications for one's own actions.

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context fully contains the expected answer, mentioning that by default one does not receive notifications about one's own activity unless explicitly targeted, and that this filter can be disabled in the 'Advanced filtering options' section of the notification settings.

Correctness: The actual output accurately describes the method to receive notifications for one's own actions and matches the expected output. Both mention accessing 'Advanced filtering options' in user settings. Minor discrepancy in wording but overall meaning aligns.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6953547654476533

Individual_scores:

AnswerRelevancy: 0.9629629629629629

Faithfulness: 0.875

ContextualPrecision: 0.5888888888888889

ContextualRecall: 1.0

CustomContextualRelevancy: 0.5

Correctness: 0.24527674083406792

Reasons:

AnswerRelevancy: The score is 0.96 because the response effectively addresses the query about disabling notifications in XWiki. However, it includes a source reference that does not directly contribute to the solution, slightly affecting the score.

Faithfulness: The score is 0.88 because the actual output incorrectly claims that XWiki allows disabling notifications about a user's own activity by default in newer versions, whereas the retrieval context states that users do not receive notifications about their own activity unless specifically targeted.

ContextualPrecision: The score is 0.59 because relevant nodes, like the second node discussing 'default watch settings' and the third node on 'filtering notifications', are ranked after an irrelevant first node that 'discusses obtaining an RSS feed for notifications'. Additionally, the fifth node, which is relevant as it 'mentions the ability for administrators to configure default notification settings', is ranked after the fourth node that 'focuses on enabling and disabling email notifications at an administrative level', which is less relevant to the user's query.

ContextualRecall: The score is 1.00 because every sentence in the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context mentions default watch settings and user profile settings but lacks details on automatic page watching and deleting custom filters.

Correctness: The actual output includes steps to manage notifications, which is additional information not in the expected output. It lacks the mention of automatic page watching and modifying notification filters from the 'Custom Filters' list, which are key elements in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8676662677198373

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 1.0
ContextualRecall: 0.4
CustomContextualRelevancy: 1.0
Correctness: 0.8893309396523577

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant to configuring OpenID Connect in XWiki, but it includes an irrelevant mention of LDAP authentication, which slightly detracts from the overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions between the actual output and the retrieval context. Great job maintaining consistency!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which contains the property '# oidc.enableUser=true', is ranked first. This directly addresses enabling user accounts on first login, while all irrelevant nodes discussing domain-based instance authentication, session cookie configuration, prerequisites, installation instructions, container authentication, and LDAP are ranked lower.

ContextualRecall: The score is 0.40 because while the retrieval context's 1st node supports the mention of the 'oidc.enableUser=true' property and its purpose, it lacks details on adding or updating properties, saving the file, restarting XWiki, and the automatic enabling of new user accounts upon first login via OpenID Connect.

CustomContextualRelevancy: The retrieval context includes the exact information needed to enable user accounts on first login by setting 'oidc.enableUser=true' in the xwiki.properties file, matching the expected output.

Correctness: The actual output accurately reflects all critical facts from the expected output, including setting the 'oidc.enableUser' property to 'true' and restarting XWiki. It misses the comment about uncommenting the line, which is minor.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.941604774656735
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 0.8
CustomContextualRelevancy: 1.0
Correctness: 0.8496286479404095

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which states 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user' and provides examples of 'oidc.user.mapping', is ranked first. This directly relates to mapping additional user profile fields, while the irrelevant nodes, such as the second node discussing scopes and response types, the third node about configuring endpoints and adding custom HTTP headers, the fourth node on group synchronization and customization of templates and listeners, and the fifth node on in-wiki configuration for OIDC authenticator and client configurations, are all ranked lower, maintaining perfect contextual precision. Great job on getting it spot on!

ContextualRecall: The score is 0.80 because the 3rd node in the retrieval context supports most of the expected output by detailing how to map user profile fields using the `oidc.user.mapping` property and providing examples. However, it lacks information about saving the `xwiki.properties` file and restarting XWiki, which affects the completeness of the retrieval.

CustomContextualRelevancy: The retrieval context contains all the information from the expected answer, including the ability to map additional user profile fields using the `oidc.user.mapping` property in the `xwiki.properties` file, with examples provided.

Correctness: The actual output accurately describes the process of mapping fields using `oidc.user.mapping` and provides a clear example similar to expected output. However, it introduces additional details about syntax and verification steps not mentioned in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.7915012734970306

Individual_scores:

AnswerRelevancy: 0.9130434782608695

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.9

Correctness: 0.5359641627213141

Reasons:

AnswerRelevancy: The score is 0.91 because the response is mostly relevant to configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes some irrelevant details about template customization and the name of the template, which slightly detract from the focus on the synchronization process.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant nodes. The first three nodes provide detailed information on configuring the `oidc.groups.claim` property, mapping synchronized groups using `oidc.groups.mapping`, and using a separator character in `oidc.groups.separator`, which are all crucial for synchronizing group claims between XWiki and an OpenID Connect provider. The irrelevant nodes, ranked fourth and fifth, only provide general information about OpenID Connect integration and configuration, which do not specifically address group synchronization.

ContextualRecall: The score is 0.40 because while nodes in retrieval context provide specific details about configuring properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, many general instructions in the expected output are not directly referenced, leading to a lower score.

CustomContextualRelevancy: The retrieval context includes details on configuring `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, which align with the expected output instructions. However, it does not explicitly mention saving the `xwiki.properties` file and restarting XWiki.

Correctness: The actual output includes relevant details like configuring group claims and mapping XWiki groups, but misses specific instructions like locating and editing the `xwiki.properties` file, saving it, and restarting XWiki. It also introduces steps not present in the expected output, such as enabling debug logging and template customization.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.616667065563859

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.6666666666666666

CustomContextualRelevancy: 0.3

Correctness: 0.5666690600498205

Reasons:

AnswerRelevancy: The score is 0.83 because the output provides relevant information on setting upload restrictions, but includes several irrelevant statements about forcing downloads and basic XWiki concepts, which do not directly address the question.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which 'provides relevant information on setting maximum attachment size and mentions the configuration of mimetype restrictions', is ranked third. Irrelevant nodes, such as the first node that 'discusses linking to external files and images, which is unrelated to setting upload restrictions on file types and sizes', should be ranked lower to improve the score.

ContextualRecall: The score is 0.67 because while several detailed steps for configuring maximum attachment size and mimetype restrictions are supported by nodes in the retrieval context, introductory and explanatory sentences, as well as step titles, are not directly found in the nodes.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specifics on configuring 'xwiki.properties' for mimetypes and setting size to 10MB.

Correctness: The actual output correctly identifies setting size limits and mimetype restrictions but uses '-attachment.validation.whitelist' instead of 'attachment.upload.allowList' and lacks precise details such as '10485760' for 10MB in bytes, and mentions additional, irrelevant configurations like '-attachment.download.forceDownload'.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.5600622301987723

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 0.5

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Correctness: 0.027040047859300458

Reasons:

AnswerRelevancy: The score is 0.83 because the response is mostly relevant to the question about denying script rights to a space administrator. However, the inclusion of a source link that does not directly address the question slightly detracts from the overall relevancy.

Faithfulness: The score is 0.50 because the actual output incorrectly describes the conditions under which scripts can be executed in XWiki. It claims that setting the 'Script' right to 'Deny' prevents script execution unless explicitly allowed, but the retrieval context specifies a priority order for rights and conditions based on the last author's permissions. Additionally, the actual output misrepresents the

default settings for script rights in the main wiki versus sub-wikis.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'programming rights and script execution' but not the specific issue of denying script rights to a space administrator. The second node lists 'various rights and their properties' but does not mention the inability to deny script rights to a space administrator. The third node talks about 'script rights' but not in the context of denying them to a space administrator. The fourth node is about 'checking access rights programmatically', which is unrelated. Finally, the fifth node discusses 'rights conversion during page migration', which is also unrelated.

ContextualRecall: The score is 1.00 because every part of the expected output is perfectly supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The expected answer states that rights implied by admin cannot be denied, which is supported by the retrieval context indicating that admin rights are not deniable and imply other rights.

Correctness: The actual output incorrectly suggests it is possible to deny the 'Script' right to space administrators, contradicting the expected output which states that rights implied by admin rights cannot be denied in XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7347222222222222

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 0.6666666666666666

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.875

CustomContextualRelevancy: 0.5

Correctness: 0.8999999999999998

Reasons:

AnswerRelevancy: The score is 0.80 because the response largely addresses the configuration question effectively, but it includes multiple irrelevant source links that do not directly contribute to solving the problem, preventing a higher score.

Faithfulness: The score is 0.67 because the actual output incorrectly states that a new page inherits rights from its parent space by default, whereas the retrieval context indicates that this inheritance can be overridden. Additionally, the actual output claims that wiki-wide rights take precedence over space-specific rights, which contradicts the retrieval context stating that page-level permissions override wiki-wide permissions.

ContextualPrecision: The score is 0.67 because the relevant nodes in the retrieval context, such as the first node mentioning 'rights inheritance and administration features' and the fourth node discussing XWiki's 'powerful and granular permission system', are ranked higher. However, the second node, which discusses the 'Nested Spaces concept' and naming conflicts, and the third node about 'string formatting and OpenID Connect properties', are ranked higher than the relevant fourth node, affecting the score. Additionally, the fifth node about 'creation of spaces and pages' is ranked above the relevant sixth node, which explains 'wiki wide rights, granular page level rights'.

ContextualRecall: The score is 0.88 because most of the expected output sentences are well-supported by nodes in the retrieval context, particularly regarding setting rights at different levels and managing access through groups. However, there is a lack of specific information in the retrieval context about adjusting rights for additional teams or spaces, which slightly impacts the score.

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and rights inheritance, which partially aligns with the expected output's steps on setting wiki-wide and space-level

rights and inheriting rights for child pages. However, it lacks specific details about creating groups for each team and setting specific rights for XWikiAllGroup and team groups.

Correctness: The actual output closely follows the expected output with precise steps for creating groups, setting wiki-wide and space-specific rights, and inheriting rights. It adds helpful details like checking priority and testing permissions, which do not impact factual correctness.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.45826911194723424

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.17385709592583004

Reasons:

AnswerRelevancy: The score is 0.91 because the response is largely relevant to the input, effectively addressing the issue with the notification bell. However, it includes repeated source links that do not directly contribute to solving the problem, slightly lowering the score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.17 because the relevant node in the retrieval context, ranked sixth, mentions editing the 'xwiki.properties' file, which is relevant to enabling notifications. However, the first five nodes are not relevant as they discuss topics like 'custom notification filters', 'alert menu and notification display', 'different ways to receive notifications', 'links and descriptions of notification features', and 'notification email settings and auto-watch modes', none of which address enabling notifications through settings or the 'xwiki.properties' file.

ContextualRecall: The score is 0.50 because while the 5th node in the retrieval context mentions editing the xwiki.properties file, which aligns with the action of enabling notifications, it does not specify setting `notifications.enabled` to `true`, leading to partial alignment with the expected output.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications through the `xwiki.properties` file or setting `notifications.enabled` to `true`.

Correctness: The actual output does not mention enabling notifications via `notifications.enabled` in `xwiki.properties` as in the expected output, and focuses on troubleshooting alert visibility instead.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3191906386071291

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.16514383164277474

Reasons:

AnswerRelevancy: The score is 0.75 because the response mostly addresses the issue of not receiving notifications, but it includes an irrelevant source link that does not pertain to the problem.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node contains 'image links and references that do not relate to notifications or their settings', which should be ranked lower. Similarly, the second and third nodes also contain 'image links and references that do not relate to notifications or their settings', and should be ranked lower as well.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any of the nodes in the retrieval context, which only contain image links and references.

CustomContextualRelevancy: The retrieval context contains image links and references unrelated to the expected answer about enabling notifications and following users or pages in XWiki.

Correctness: Actual Output lacks key details like setting `notifications.enabled` to `true` and following users/pages, deviating from Expected Output's comprehensive guide.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.3423377803423145

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.23584486387206885

Reasons:

AnswerRelevancy: The score is 0.82 because the response is mostly relevant to the question about enabling notifications for one's own actions, but it includes source links that do not directly address the issue, which slightly detracts from the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'customization of page templates and internationalization', which is unrelated to notifications or user activity settings. The second node is about 'adminstrating wiki features' and does not mention notifications or user activity settings. The third node covers 'importing documents, sharing pages, and other unrelated features', without any mention of notification settings or user activity. Therefore, no relevant nodes are ranked higher than irrelevant nodes.

ContextualRecall: The score is 0.00 because none of the expected output sentences can be linked to any nodes in the retrieval context, as the context lacks information on notifications, user activity, or related settings.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user profile options, which are present in the expected output.

Correctness: The actual output contains some relevant information about enabling notifications but misses key facts such as the default setting of not receiving notifications unless specified and the option to disable the filter in 'Erweiterten Filteroptionen'. It includes additional steps and details not present in the expected output, leading to discrepancies.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.31703183141653046

Individual_scores:

AnswerRelevancy: 0.7916666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.11052432183251618

Reasons:

AnswerRelevancy: The score is 0.79 because the output contains some relevant information about disabling notifications, but it includes multiple irrelevant links to documentation pages that do not directly address the question about disabling notifications.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'attaching files', the second node is about 'moving attachments', the third node covers 'displaying various elements', the fourth node focuses on 'administrating wiki features', and the fifth node talks about 'importing office documents'. None of these topics relate to managing or deactivating notification settings, which is the focus of the input.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about automatic page following, notification settings, or custom filters as mentioned in the expected output.

Correctness: The actual output discusses disabling notifications, not the automatic following of pages after edits as in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.34389009222995476

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 0.8571428571428571

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Correctness: 0.02841991845909374

Reasons:

AnswerRelevancy: The score is 0.78 because the response partially addresses the question about configuring XWiki for automatic user account activation upon first login. However, it includes irrelevant information about deactivating OpenID Connect authentication and using a URL parameter, which does not pertain to the main issue.

Faithfulness: The score is 0.86 because the actual output incorrectly suggests setting the session cookie on a top-level domain, which contradicts the retrieval context that states this approach will not work if the XWiki instance has very different domains for its wikis.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication with a domain-based instance and session cookie configuration', which does not relate to automatic user account activation. The second node provides 'general information about OpenID Connect authentication' but lacks specifics on the `oidc.enableUser` property. The third node talks about 'OpenID Connect authenticator compatibility' without addressing automatic activation. Lastly, the fourth node lists 'various OpenID Connect endpoints' without mentioning the `oidc.enableUser` property necessary for user activation.

ContextualRecall: The score is 0.40 because while the 4th node in the retrieval context supports the instructions to update the xwiki.properties file and set 'oidc.enableUser' to true (as mentioned in sentences 1 and 3 of the expected output), it lacks direct references to automatically activating user accounts, saving the file, and restarting the XWiki instance (as mentioned in sentences 4, 5, and 6).

CustomContextualRelevancy: The retrieval context does not mention the property 'oidc.enableUser' or its configuration in 'xwiki.properties' for automatic user account activation.

Correctness: The actual output is unrelated to the expected output, discussing session cookies and disabling authentication instead of enabling user accounts on first login via OpenID Connect by modifying xwiki.properties.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.40428086871858343

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.42568521231150036

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about customizing user profile fields from an OpenID Connect provider in XWiki. Great job!

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect endpoints' but lacks information on 'user profile field mapping or `xwiki.properties`'. Similarly, the second node focuses on 'container authentication and other methods like LDAP and Active Directory', without addressing 'user profile field customization or `xwiki.properties`'. The third node provides a 'general description of OpenID Connect authentication' but does not cover 'user profile field mapping or `xwiki.properties`'. The fourth node offers 'metadata about the OpenID Connect Authenticator extension', yet it omits 'user profile field mapping or `xwiki.properties`'. Lastly, the fifth node lists 'dependencies for the OpenID Connect Authenticator extension', failing to mention 'user profile field mapping or `xwiki.properties`'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output have corresponding information in any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about customizing user profile fields using `xwiki.properties` or `oidc.user.mapping`, which are key elements in the expected output.

Correctness: The actual output mentions using `xwiki.cfg` instead of `xwiki.properties` for configuration and does not provide the specific mapping example with `oidc.user.mapping` as in the expected output. It includes additional configuration details not directly related to mapping user profile fields.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_017_result.json

Overall_score: 0.37818237412266864
Individual_scores:
AnswerRelevancy: 0.9047619047619048
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.07692307692307693
CustomContextualRelevancy: 0.0
Correctness: 0.28740926305103043
Reasons:

AnswerRelevancy: The score is 0.90 because the answer is mostly relevant and provides a good explanation of the process for setting up group synchronization between XWiki and an OpenID Connect provider. However, the inclusion of source references that do not directly address the setup process slightly detracts from the overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. Each node fails to mention 'group synchronization' or the 'specific properties needed for it', which are crucial for the input query. For example, the first node discusses 'various OpenID Connect endpoints and configurations' without addressing group synchronization, and the second node talks about 'authentication methods and extensions' without covering the necessary configuration steps for group synchronization.

ContextualRecall: The score is 0.08 because the retrieval context only supports the mention of 'oidc.groups.claim' related to OpenID Connect configuration (node 1 in retrieval context), while it lacks information on other key aspects such as group synchronization setup, file handling, and specific property configurations described in the expected output sentences.

CustomContextualRelevancy: The retrieval context lacks any mention of group synchronization, xwiki.properties configuration, or specific properties like oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping, which are detailed in the expected output.

Correctness: The actual output includes many unsupported facts not in the expected output such as installing the OpenID Connect Authenticator Extension and configuring various endpoints. It misses key facts like configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping'.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_018_result.json

Overall_score: 0.6062713876700743
Individual_scores:
AnswerRelevancy: 0.8333333333333334
Faithfulness: 1.0
ContextualPrecision: 0.325
ContextualRecall: 0.36363636363636365
CustomContextualRelevancy: 0.5
Correctness: 0.6156586290507481
Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the question about setting upload restrictions for images or PDFs, but includes irrelevant links about exports, XWiki syntax, and document lifecycle, which are not directly related to the query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the first three nodes in the retrieval context are unrelated to configuring attachment restrictions, discussing topics like 'trusted domains and PDF export look & feel', 'ODT export and PDF export applications', and 'various features and macros'. These irrelevant nodes are ranked higher than the relevant nodes, which mention 'Size Limit' and 'Mimetype Restriction' and provide detailed steps on setting the maximum size of an attachment. This misordering affects the score, as relevant information is not prioritized.

ContextualRecall: The score is 0.36 because while some detailed steps about configuring maximum upload size are supported by the 5th node in the retrieval context, key instructions about MIME type configuration and the overall purpose of limiting uploads to images or PDFs up to 10 MB are not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and MIME type restrictions but lacks specific details on configuring xwiki.properties and setting size to 10 MB.

Correctness: The actual output correctly addresses setting the upload size to 10 MB and mentions editing XWikiPreferences, but lacks specific instructions on configuring MIME types in xwiki.properties and suggests using 'Attachment Validation Application' without details.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.3202407385369275

Individual_scores:

AnswerRelevancy: 0.7647058823529411

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.15673854886862415

Reasons:

AnswerRelevancy: The score is 0.76 because the response partially addresses the question about denying script rights to a domain administrator, but includes irrelevant details such as source links and unrelated advice, which prevent it from achieving a higher score.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question. The first node discusses 'form validation' without mentioning script rights or administrative permissions. The second node covers 'template providers and page creation,' which is unrelated to the topic of script rights. The third node talks about a 'custom right check class and configuration' but does not address denying script rights to an area administrator. Therefore, irrelevant nodes are ranked higher, resulting in a low score.

ContextualRecall: The score is 0.00 because the expected output sentence about rights and permissions in XWiki is not supported by any information in the nodes in retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about XWiki not supporting denial of rights implied by admin rights.

Correctness: The actual output provides methods to deny script rights, which contradicts the expected output stating it is unsupported. Also, the actual output lacks the explicit statement that admin-implied rights cannot be denied.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.41664856687228863

Individual_scores:

AnswerRelevancy: 0.7916666666666666

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.125

CustomContextualRelevancy: 0.0

Correctness: 0.3332247345670652

Reasons:

AnswerRelevancy: The score is 0.79 because the response largely addresses the question about configuring access and editing permissions in XWiki, but it includes repeated irrelevant source links that do not directly relate to the specific configuration task, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.25 because the relevant node in retrieval context, which discusses 'managing users and groups in XWiki', is ranked fourth. This node should be ranked higher than the first node, which discusses 'XWiki's database support and programming features', the second node, which is about 'notifications and application details in XWiki', the third node, which explains 'the use of nested pages and subwikis', and the fifth node, which covers 'group synchronization and customization', as these are not relevant to configuring access control for teams.

ContextualRecall: The score is 0.12 because only sentence 1 in the expected output is supported by node 4 in the retrieval context, which mentions managing users and groups. The rest of the sentences in the expected output have no corresponding information in any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to access control steps or rights management as outlined in the expected output.

Correctness: The actual output partially follows the expected output with group creation and access control, but lacks specific steps like setting wiki-wide rights and inheriting rights for subpages. It introduces extraneous elements like navigation and group synchronization not present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.44688953989814256

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Correctness: 0.28660039728359205

Reasons:

AnswerRelevancy: The score is 0.89 because the response mostly addresses the issue of the notification bell not appearing, but it includes irrelevant information about XWiki syntax documentation and incorrectly suggests the extension is not installed.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'custom notification filters and their behavior', which is unrelated to 'enabling notifications in the wiki settings'. The second node lists 'various notification-related settings and customization options', but does not mention 'enabling notifications in the wiki settings'. The third node describes 'notification preferences and macros', but does not provide information on 'enabling notifications in the wiki settings'. The fourth node provides a description of the 'notifications application and its features', but does not address 'enabling notifications in the wiki settings'. The fifth node discusses 'obtaining an RSS feed for notifications and settings', but does not mention 'enabling notifications in the wiki settings'.

ContextualRecall: The score is 0.50 because sentence 1 is supported by node 5 in the retrieval context, which discusses activating notifications. However, sentence 2 lacks support as no nodes mention setting 'notifications.enabled' in 'xwiki.properties'.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the 'xwiki.properties' file or the 'notifications.enabled' property, which is the main information in the expected answer.

Correctness: The actual output addresses the issue of missing notifications but does not directly reference enabling notifications in 'xwiki.properties', unlike the expected output. It provides alternative reasons for the problem, which are not mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.692174976758312

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Correctness: 0.26971652721653844

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the issue of not receiving notifications, but includes irrelevant source links that do not directly relate to solving the problem.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than irrelevant nodes. The first node discusses 'Enable and disable the extension' and 'Enabling notifications for all users by default', which are directly related to ensuring notifications are activated. The second node aligns with the suggestion to follow users to receive notifications. Irrelevant nodes, such as the third node discussing 'getting a notification RSS feed and notification settings' and the fourth node listing 'various notification preferences macros', are correctly ranked lower as they do not directly address enabling notifications or following users. Great job on getting everything in the right order!

ContextualRecall: The score is 0.75 because while the nodes in the retrieval context cover most aspects of enabling notifications and following users or pages (as seen in sentences 1, 3, 4, 5, and 6 of the expected output), they lack specific details about 'xwiki.properties' and the 'réseau' tab mentioned in sentences 2 and 7.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' to 'true' in 'xwiki.properties' and specific steps for following pages or users.

Correctness: The actual output discusses notification preferences and possible issues but lacks specific instructions on enabling notifications in xwiki.properties and does not mention following pages or users as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8784389240736262

Individual_scores:

AnswerRelevancy: 0.6

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Correctness: 0.7706335444417578

Reasons:

AnswerRelevancy: The score is 0.60 because the output includes source links that do not directly address the question about enabling notifications for one's own actions, which makes it partially relevant but not fully aligned with the input query.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because the relevant node in retrieval context, ranked first, directly addresses the input question by stating 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' ensuring that all irrelevant nodes are ranked lower.

ContextualRecall: The score is 1.00 because all sentences in the expected output are fully supported by the nodes in the retrieval context, with no unsupportive reasons identified. Great job!

CustomContextualRelevancy: The retrieval context confirms that by default, users do not receive notifications for their own actions, and this can be changed by adjusting the 'Own event filter'. However, it does not explicitly mention the 'Options de filtrage avancées' section in the user profile settings.

Correctness: The actual output correctly describes disabling a filter to receive notifications on personal activities, aligning with expected output. However, it lacks mention of the 'Options de filtrage avancées' section for modifying these settings.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7588655493199252

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 1.0
ContextualRecall: 0.75
CustomContextualRelevancy: 0.8
Correctness: 0.15703944976570502

Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses how to disable notifications, but includes irrelevant links that do not directly help with the question.

Faithfulness: The score is 1.00 because there are no contradictions. Everything aligns perfectly with the retrieval context. Great job!

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant node. The first three nodes provide information about 'Choosing the default auto-watched mode', 'Notifications Auto Watch Preferences', and 'Auto Watch', which are directly related to managing automatic notifications. The irrelevant node, ranked fourth, discusses 'Get a notification RSS feed' and 'Notifications Macro', which are not directly related to the input query about disabling automatic notifications.

ContextualRecall: The score is 0.75 because the retrieval context supports most of the expected output, particularly regarding automatic page watching and notification settings (as mentioned in the 2nd node in retrieval context). However, it lacks information about being notified of changes to your work, which affects the overall score.

CustomContextualRelevancy: The retrieval context includes information about auto-watched pages and custom notification filters, aligning with the expected answer's mention of automatic page following and notification settings. However, it lacks explicit details about modifying these settings in the user profile.

Correctness: The actual output discusses how to manage notifications but does not mention the specific features of 'automatic page following' or 'custom filters' as in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.4199754690107147
Individual_scores:
AnswerRelevancy: 0.9285714285714286
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.4
CustomContextualRelevancy: 0.0
Correctness: 0.1912813854928596

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to configuring OpenID Connect in XWiki, with only a minor irrelevant mention of the LDAP Authenticator.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'domain-based authentication and session cookies', which is unrelated to enabling user accounts automatically with OpenID Connect. The second node mentions 'container authentication and LDAP', not addressing OpenID Connect or automatic user activation. The third node, although discussing OpenID Connect, focuses on 'authentication and synchronization', not on enabling user accounts automatically. The fourth node lists 'dependencies for the OpenID Connect extension', which is not relevant to the required configuration. Lastly, the fifth node is about 'LDAP and Active Directory integration', which does not pertain to the input query.

ContextualRecall: The score is 0.40 because while the 3rd node in the retrieval context mentions 'OpenID Connect' and 'xwiki.properties', which supports the general context of adjusting the file for OpenID Connect, it does not specifically cover adding or updating the property, saving the file, or ensuring automatic activation of user accounts as described in the expected output.

CustomContextualRelevancy: The retrieval context lacks any mention of modifying the xwiki.properties file or setting the oidc.enableUser property to true for automatic user account activation.

Correctness: The actual output incorrectly references `xwiki.authentication.createuser` in `xwiki.cfg` instead of `oidc.enableUser` in `xwiki.properties`, and suggests creating an empty user, which does not match the expected procedure of enabling a user.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.4705176599010253

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.2

Correctness: 0.3283691173008884

Reasons:

AnswerRelevancy: The score is 0.89 because the answer is mostly relevant to the question about customizing user profile fields in XWiki, but it includes unnecessary source links that do not directly address the question, preventing a perfect score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect endpoints,' but does not mention 'user profile field mapping or customization.' Similarly, the second node is about 'container authentication and LDAP,' with a brief mention of OpenID Connect, but it does not address 'user profile field mapping or customization.' The third node provides 'general information about the OpenID Connect Authenticator extension,' without discussing 'user profile field mapping or customization.' The fourth node lists 'dependencies for the OpenID Connect Authenticator extension,' which is unrelated to 'user profile field mapping or customization.' Finally, the fifth node describes 'configuration guides and limitations for OpenID Connect,' but does not mention 'user profile field mapping or customization.'

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention 'xwiki.properties' and 'oidc.user.mapping', which supports the customization of user profile fields from OpenID Connect as in sentence 1, they lack details on specific mapping examples, property replacements, and the process of saving and restarting XWiki as described in sentences 2 to 5.

CustomContextualRelevancy: The retrieval context mentions OpenID Connect but lacks specific details about customizing user profile fields or using `oidc.user.mapping` in `xwiki.properties`, which are key elements in the expected output.

Correctness: Actual output uses `xwiki.cfg` instead of `xwiki.properties` for mapping, and `oidc.attribute.mappings` instead of `oidc.user.mapping`. It introduces steps not present in expected output and lacks specific mapping examples using `\${oidc.user.subject}`.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5023124773615483

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.3076923076923077

CustomContextualRelevancy: 0.4

Correctness: 0.2585635088579338

Reasons:

AnswerRelevancy: The score is 0.71 because the output contains some relevant information about configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes several irrelevant statements, such as source links, file names, and unrelated configuration details, which do not directly address the process, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which provides information about group synchronization and mentions the use of `oidc.groups.mapping` in the xwiki.properties file, is ranked third. However, it is preceded by irrelevant nodes that discuss authentication management and endpoints for OpenID Connect (first node), various authentication methods and extensions (second node), and general information about OpenID Connect authentication (fourth node), which do not address group synchronization or configuration.

ContextualRecall: The score is 0.31 because while the 3rd node in the retrieval context supports the configuration of group synchronization in XWiki, it lacks specific details such as locating and editing the 'xwiki.properties' file, providing examples, and instructions on saving and restarting XWiki, which are crucial for a complete understanding.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using oidc.groups.mapping, but lacks details on configuring oidc.groups.claim and oidc.groups.separator as specified in the expected output.

Correctness: The output introduces `oidc.userinfoclaims` and `xwiki.cfg` which are incorrect, and lacks details on `oidc.groups.claim`, `oidc.groups.separator`, and saving changes in `xwiki.properties`, as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.32674146931716397

Individual_scores:

AnswerRelevancy: 0.6875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.2729488159029837

Reasons:

AnswerRelevancy: The score is 0.69 because the response partially addresses the question about file upload restrictions, but includes multiple irrelevant links that do not specifically relate to the query about limiting uploads to image or PDF files of a certain size.

Faithfulness: The score is 1.00 because there are no contradictions, indicating a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'language preferences and translations in XWiki', which is unrelated to file upload restrictions. The second node is about 'various XWiki applications and their functionalities, such as mail configuration and PDF export', which do not address file upload restrictions. The third node provides 'technical details about CSS and XSL transformations for PDFs', which is not relevant to configuring file upload restrictions. The fourth node lists 'various XWiki features and macros, such as PDF Viewer and Office Document Viewer', which do not relate to file upload restrictions. The fifth node contains 'image links and references', which are unrelated to the topic of file upload restrictions. The sixth node also includes 'image links and references', which do not provide any information on file upload restrictions.

ContextualRecall: The score is 0.00 because none of the expected output sentences can be attributed to any nodes in the retrieval context, as the context lacks information on file size restrictions, MIME type configurations, and specific instructions related to the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information related to configuring MIME types or setting maximum attachment size as described in the expected output.

Correctness: Actual output lacks MIME type configuration and does not mention editing xwiki.properties as in expected output. It only addresses file size setting in a generic manner.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.5375137728311569

Individual_scores:

AnswerRelevancy: 0.7

Faithfulness: 0.8333333333333334

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Correctness: 0.1917493036536082

Reasons:

AnswerRelevancy: The score is 0.70 because the response partially addresses the question about refusing script rights, but includes irrelevant links to XWiki documentation that do not directly answer the query.

Faithfulness: The score is 0.83 because the actual output incorrectly claims that the 'SCRIPT' right can be denied specifically to an administrator, whereas the retrieval context states that the 'SCRIPT' right is deniable without specifying any particular restrictions on administrators.

ContextualPrecision: The score is 0.50 because the second node in the retrieval context is relevant as it discusses 'programming and script rights,' which is pertinent to the question about denying script rights. However, the first node, which is about 'language preferences and translations in XWiki,' is irrelevant and should be ranked lower. Additionally, the third node, which lists 'various rights and their default states,' is not directly related to denying script rights and should also be ranked lower than the relevant node.

ContextualRecall: The score is 1.00 because the expected output is perfectly supported by the information in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context does not mention that administrator rights cannot be refused, which is the key information in the expected output.

Correctness: The actual output incorrectly suggests methods to refuse admin rights, contradicting the expected output which states it's not possible in XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.3547757478995622

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.3508767096195956

Reasons:

AnswerRelevancy: The score is 0.78 because the response largely addresses the question about configuring spaces and roles in XWiki, but the repeated inclusion of source links, which are not directly relevant to the configuration instructions, slightly detracts from the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'a specific XML page structure in XWiki', which does not relate to configuring access rights for teams. The second node talks about 'modifying resources using XML and HTTP methods', which is unrelated to setting access permissions in XWiki. The third node is about 'adding and managing comments in XWiki', which does not pertain to configuring team access to spaces. The fourth node explains 'the emulation of spaces and pages in XWiki', but does not provide information on setting access rights for teams. The fifth node discusses 'migration actions and changes in XWiki', which are not relevant to configuring team access to spaces. Lastly, the sixth node is about 'configuration display scopes in XWiki', which is not directly related to setting up access permissions for teams.

ContextualRecall: The score is 0.00 because none of the expected output sentences are supported by the nodes in the retrieval context, as they lack any mention of the specific steps, group creation, rights setting, or inheritance described in the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting rights at wiki or space levels, or inheriting rights for child pages as mentioned in the expected output.

Correctness: Actual Output mentions similar goals but lacks specific steps like creating groups, setting wiki-level rights, and inheriting rights for child pages as in Expected Output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7025804812124959

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.7

Correctness: 0.523419395211484

Reasons:

AnswerRelevancy: The score is 0.78 because the response partially addresses the question about authenticating users with access tokens, but includes multiple irrelevant statements providing source links that do not directly contribute to the answer.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.50 because the relevant nodes in the retrieval context are not consistently ranked higher than irrelevant nodes. The first node, ranked 1, discusses 'LDAP authentication settings, which are unrelated to token-based authentication or JWTs,' and should be ranked lower. Similarly, the third node, ranked 3, 'discusses OpenID Connect scopes and response types, which are not directly related to the use of access tokens or JWTs for authentication,' and should also be ranked lower. However, the second node, ranked 2, 'provides information on configuring authorized applications, which is relevant to using access tokens for authentication,' and the fourth node, ranked 4, 'describes the use of JWTs for authentication, which directly relates to the expected output about authenticating users with access tokens,' are correctly identified as relevant, contributing to the current score.

ContextualRecall: The score is 0.71 because while nodes in the retrieval context support several key aspects of the expected output, such as authenticating users with JWTs and configuring authorized applications (nodes 2 and 4), they lack information on generating JWT tokens with specific claims and including them in the 'Authorization' header.

CustomContextualRelevancy: The retrieval context contains information about enabling the token authenticator and configuring authorized applications, but lacks details on JWT claims, signing key requirements, and the specific use of the 'Authorization' header.

Correctness: The actual output correctly mentions the use of the LLM Application Authenticator and JWT for authentication, aligning with the expected output. However, it inaccurately includes OpenID Connect and omits details on token claims and Ed25519 key, which affects factual accuracy.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.0

Individual_scores:

AnswerRelevancy: 0

Faithfulness: 0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0

Reasons:

AnswerRelevancy: Failed to measure after 3 attempts: Request failed with status code 500:

{"error":{"message":"Invalid 'messages[0].content': string too long. Expected a string with maximum length 1048576, but got a string with length 1251217 instead.","type":null,"param":null,"code":"400"}}

Faithfulness: Failed to measure after 3 attempts: Request failed with status code 500:

{"error":{"message":"Invalid 'messages[0].content': string too long. Expected a string with maximum length 1048576, but got a string with length 1251500 instead.","type":null,"param":null,"code":"400"}}

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node is about 'Model Validation' and does not mention authentication or access tokens, the second node discusses 'Attachment Validation' which is unrelated to token-based authentication, and the third node provides information about the 'Admin UI for Model Validation', which is not relevant to authenticating users with access tokens.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.
CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are central to the expected output.
Correctness: Failed to measure after 3 attempts: Error code: 400 - {'error': {'message': 'Invalid 'messages[0].content': string too long. Expected a string with maximum length 1048576, but got a string with length 1253153 instead.', 'type': 'invalid_request_error', 'param': 'messages[0].content', 'code': 'string_above_max_length'}}
Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_033_result.json

Overall_score: 0.23851461715307407

Individual_scores:

AnswerRelevancy: 0.4

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Correctness: 0.031087702918444416

Reasons:

AnswerRelevancy: The score is 0.40 because the response includes some relevant information about user authentication, but it is overshadowed by numerous irrelevant statements that do not directly address the question about using access tokens for authentication.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'upgrading XWiki' and does not mention 'authentication or access tokens'. The second node focuses on 'user and rights management in XWiki' without addressing 'authentication with access tokens'. The third node talks about 'rights management and user profiles' but lacks information on 'authenticating users with access tokens'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, which are key elements in the expected output.

Correctness: The actual output does not mention JWT or token-based authentication and instead discusses user rights and profiles, which are unrelated to the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Summarization Results

Model: AI.Models.GPT4o

File: summ_001_result.json

Score: 0.4

Reason: The score is 0.40 because the summary includes extra information not present in the original text, such as the ability to print pages and the creation of multiple wikis. Additionally, the summary fails

to address several questions that the original text can answer, indicating a lack of coverage and accuracy.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.6

Reason: The score is 0.60 because the summary includes extra information not present in the original text, such as a tabular view summarizing rights, which may not align with the original content. Additionally, the summary fails to address specific questions that the original text can answer, indicating a lack of completeness and accuracy.

Score_breakdown:

Alignment: 0.8888888888888888

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.3333333333333333

Reason: The score is 0.33 because the summary includes a significant amount of extra information not present in the original text, such as details about managing user groups, implications of denials, and additional resources, which suggests a lack of alignment with the original content.

Score_breakdown:

Alignment: 0.3333333333333333

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.2857142857142857

Reason: The score is 0.29 because the summary contains a significant contradiction regarding the use of the Distribution Wizard, includes multiple pieces of extra information not present in the original text, and fails to answer a question that the original text can address. These issues indicate a low-quality summarization.

Score_breakdown:

Alignment: 0.2857142857142857

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.5

Reason: The score is 0.50 because the summary contains a contradiction regarding the requirement of application installation for watchlist modifications, which is not mentioned in the original text. Additionally, it includes extra information about editing a specific page for RSS feeds that is not present

in the original text. Furthermore, the summary fails to answer a question about the availability of a feed for all modified pages, which the original text can address.

Score_breakdown:

Alignment: 0.5

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.7142857142857143

Reason: The score is 0.71 because the summary includes extra information not present in the original text, such as the need for credentials from an administrator and steps for creating a page, which affects its alignment with the original content.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.8333333333333334

Reason: The score is 0.83 because the summary includes extra information about a 'Version Control page' in XWiki that is not mentioned in the original text. However, there are no contradictions, indicating a generally accurate summary.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.6

Reason: The score is 0.60 because the summary includes extra information not present in the original text and fails to answer specific questions that the original text can address. However, there are no contradictions, indicating some level of accuracy.

Score_breakdown:

Alignment: 0.8

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.8571428571428571

Reason: The score is 0.86 because the summary includes extra information about XWiki and its reference pages that is not mentioned in the original text. However, the absence of contradictions suggests a generally accurate summary.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 1.0
Input_language: en
Summary_language: en

File: summ_010_result.json

Score: 0.8
Reason: The score is 0.80 because the summary includes extra information about examples of applications like blogs or task managers, which were not specified in the original text. However, there are no contradictions, indicating a generally accurate summary.
Score_breakdown:
Alignment: 0.8
Coverage: 1.0
Input_language: en
Summary_language: en

Model: AI.Models.GPT4o-mini

File: summ_001_result.json

Score: 0.625
Reason: The score is 0.62 because the summary includes incorrect information about the typical use of panels, adds details about wikis and sections not present in the original text, and fails to answer a question about email notifications that the original text can address.
Score_breakdown:
Alignment: 0.625
Coverage: 0.8
Input_language: en
Summary_language: en

File: summ_002_result.json

Score: 0.6
Reason: The score is 0.60 because the summary contains a contradiction regarding where Programming Right can be granted, includes extra information about priority orders and permission precedence not mentioned in the original text, and fails to answer questions about specific rights and permissions that the original text can address.
Score_breakdown:
Alignment: 0.7272727272727273
Coverage: 0.6
Input_language: en
Summary_language: en

File: summ_003_result.json

Score: 0.45454545454545453
Reason: The score is 0.45 because the summary includes several pieces of extra information not present in the original text, such as the creation of groups for easier management and specific instructions for setting rights. Additionally, the summary fails to answer a question that the original text can address, indicating a lack of completeness and accuracy.
Score_breakdown:
Alignment: 0.45454545454545453

Coverage: 0.8
Input_language: en
Summary_language: en

File: summ_004_result.json

Score: 0.4444444444444444

Reason: The score is 0.44 because the summary contains multiple contradictions with the original text, such as incorrect methods for upgrading distribution and misrepresentations of alternative methods for starting fresh. Additionally, there is extra information in the summary not present in the original text, like troubleshooting tips, and it fails to answer questions that the original text can, such as migration paths for upgrading XWiki versions.

Score_breakdown:

Alignment: 0.4444444444444444

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains a contradiction regarding the default RSS feeds and the Watchlist Feature, and it fails to answer questions about creating new RSS feeds and receiving RSS feeds for blog posts.

Score_breakdown:

Alignment: 0.75

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.3333333333333333

Reason: The score is 0.33 because the summary contains multiple contradictions and extra information not present in the original text, such as the scope of the Getting Started guide and details about user actions. Additionally, the summary fails to address certain questions that the original text can answer, indicating a lack of coverage and accuracy.

Score_breakdown:

Alignment: 0.3333333333333333

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.8333333333333334

Reason: The score is 0.83 because the summary includes extra information about a 'Version Control page on XWiki.org' that is not mentioned in the original text. However, there are no contradictions, indicating a generally accurate summary.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 1.0

Input_language: en
Summary_language: en

File: summ_008_result.json

Score: 0.7142857142857143

Reason: The score is 0.71 because the summary includes extra information not present in the original text, such as accompanying images for configurations and guidance for creating a basic application. However, there are no contradictions, indicating a generally accurate representation of the original content.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 1.0

Reason: The score is 1.00 because the summary perfectly aligns with the original text, with no contradictions or unnecessary additions.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary includes extra information about specific applications like blogs, task managers, FAQs, and product management tools that are not mentioned in the original text, which slightly reduces its accuracy.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.claude3_5_sonnet

File: summ_001_result.json

Score: 0.7142857142857143

Reason: The score is 0.71 because the summary includes extra information not present in the original text, such as details about starting with a single wiki and customizing panel layouts. Additionally, the summary fails to address a question that the original text can answer regarding exporting a page in PDF format. However, there are no contradictions, indicating a generally accurate representation of the original content.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 0.8

Input_language: en
Summary_language: en

File: summ_002_result.json

Score: 0.0

Reason: The score is 0.00 because the summary contains multiple contradictions with the original text, such as incorrectly categorizing rights and misrepresenting the permissions system. Additionally, the summary fails to answer specific questions that the original text addresses, indicating a significant lack of alignment between the summary and the original content.

Score_breakdown:

Alignment: 0.0

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.36363636363636365

Reason: The score is 0.36 because the summary contains significant contradictions and adds extra information not present in the original text, leading to a low-quality summarization.

Score_breakdown:

Alignment: 0.36363636363636365

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.5454545454545454

Reason: The score is 0.55 because the summary contains a significant contradiction regarding the recommended downgrading method and includes several pieces of extra information not present in the original text. These issues indicate a moderate level of accuracy and completeness in the summary.

Score_breakdown:

Alignment: 0.5454545454545454

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.75

Reason: The score is 0.75 because the summary contains a contradiction regarding the functionality of the Watchlist Feature and includes extra information about Main.WebRss and custom feeds that is not present in the original text.

Score_breakdown:

Alignment: 0.75

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.4166666666666667

Reason: The score is 0.42 because the summary includes extra information not present in the original text, such as details about login credentials, user permissions, developer access, and documentation links. Additionally, the summary fails to answer whether XWiki is described as a second generation wiki, which the original text can answer. These issues indicate a lack of alignment and completeness between the summary and the original text.

Score_breakdown:

Alignment: 0.4166666666666667

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.8888888888888888

Reason: The score is 0.89 because the summary includes extra information about XWiki.org's Version Control page that is not mentioned in the original text. However, there are no contradictions, indicating a generally accurate summary with minor additions.

Score_breakdown:

Alignment: 0.8888888888888888

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.8

Reason: The score is 0.80 because the summary includes extra information not present in the original text and fails to answer a question that the original text addresses. However, there are no contradictions, indicating a generally accurate representation of the original content.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.8571428571428571

Reason: The score is 0.86 because the summary includes extra information comparing wikis to traditional content management systems, which is not mentioned in the original text. However, there are no contradictions, indicating a generally accurate summary.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.6

Reason: The score is 0.60 because the summary includes extra information not present in the original text, such as custom applications like blogs or task managers and users customizing the platform while maintaining core functionality. However, there are no contradictions, which indicates some level of accuracy.

Score_breakdown:

Alignment: 0.6

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.command-r_35B_Q4

File: summ_001_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains contradictory information about the customization of panels, includes extra information about printing pages not mentioned in the original text, and fails to answer several questions that the original text can address, such as editing with a WYSIWYG editor, creating multiple wikis, and exporting pages in PDF format.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains contradictions regarding the administration rights and editing permissions, which affects the accuracy. Additionally, the summary fails to address specific questions about default edit rights and delete rights, indicating incomplete coverage of the original text.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.1111111111111111

Reason: The score is 0.11 because the summary contains significant contradictions and introduces numerous pieces of extra information not present in the original text. Additionally, it fails to answer questions that the original text can address, indicating a poor alignment with the original content.

Score_breakdown:

Alignment: 0.1111111111111111

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.3333333333333333

Reason: The score is 0.33 because the summary contains significant contradictions and includes extra information not found in the original text. It also fails to answer questions that the original text can address, indicating a lack of alignment and completeness.

Score_breakdown:

Alignment: 0.3333333333333333

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.6

Reason: The score is 0.60 because the summary fails to address specific questions that the original text can answer, indicating that some important details are missing from the summary. However, there are no contradictions or extra information, which suggests that the summary is generally accurate but lacks completeness.

Score_breakdown:

Alignment: 1.0

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.6

Reason: The score is 0.60 because the summary includes extra information not present in the original text, such as users learning to create pages and documentation links, and it fails to answer questions about specific features like dedicated wiki documentation and pre-installed applications, indicating a moderate level of alignment with the original text.

Score_breakdown:

Alignment: 0.6

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains significant inaccuracies, such as incorrect descriptions of features related to version comparison and reverting changes. Additionally, it introduces information not present in the original text, like a 'Version Control' page, and fails to address questions that the original text can answer, indicating a lack of coverage and precision.

Score_breakdown:

Alignment: 0.4

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains a contradiction regarding how global rights are accessed, includes extra information about screenshots not present in the original text, and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary contains a contradiction regarding who can modify wiki pages and includes extra information about XWiki and references not present in the original text. Additionally, it fails to address whether most wikis include features like access rights management, which the original text can answer.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.625

Reason: The score is 0.62 because the summary includes extra information not present in the original text, such as specific features and examples of XWiki's use cases, which were not detailed in the original text. This indicates a moderate level of alignment between the summary and the original text, but the inclusion of additional information affects the accuracy.

Score_breakdown:

Alignment: 0.625

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.gemma2_9B_Q4

File: summ_001_result.json

Score: 0.375

Reason: The score is 0.38 because the summary contains significant contradictions regarding the content of the tabbed area and panel customization, includes extra information not present in the original text such as details about XWiki as a platform and features like creating child pages, and fails to answer questions about editing and exporting capabilities that the original text covers.

Score_breakdown:

Alignment: 0.375

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains significant contradictions with the original text, such as incorrect claims about the 'createwiki' right and default permissions in XWiki. Additionally, the summary includes extra information not present in the original text, like details about register permission and inheritance rules. Furthermore, the summary fails to answer specific questions that the original text addresses, such as the default status of certain rights and the capabilities of the script right.

Score_breakdown:

Alignment: 0.7058823529411765

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.375

Reason: The score is 0.38 because the summary includes multiple pieces of extra information not present in the original text, such as details about permission levels, types, and behaviors. Additionally, the summary fails to address a question that the original text can answer regarding permission overrides in XWiki.

Score_breakdown:

Alignment: 0.375

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.5714285714285714

Reason: The score is 0.57 because the summary contains contradictory information about the Flavor Upgrade process and includes extra details not present in the original text, such as specific commands for upgrading and potential issues with database schema compatibility. Additionally, the summary fails to address whether XWiki provides migration paths for upgrades, which the original text can answer.

Score_breakdown:

Alignment: 0.5714285714285714

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary includes extra information not found in the original text, such as the requirement of a Watchlist application for the Watchlist RSS feed and specific details about the 'Main.WebRss' page and example implementations. However, there are no contradictions, indicating a generally accurate summary with some unnecessary additions.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary includes extra information not present in the original text, such as details about creating, editing, and viewing page history, and encouragement to proceed to Step 1. Additionally, it fails to answer a question about the existence of a dedicated wiki for XWiki Rendering documentation, which the original text can answer.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.7777777777777778

Reason: The score is 0.78 because the summary contains a minor contradiction regarding the location of the 'History' tab and includes extra information about a tutorial step not present in the original text. Despite these issues, the summary is generally accurate and informative.

Score_breakdown:

Alignment: 0.7777777777777778

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.4

Reason: The score is 0.40 because the summary includes extra information not present in the original text, such as a guide on setting user rights, screenshots illustrating configurations, and creating a basic XWiki application. Additionally, the summary fails to answer a question that the original text can, indicating a lack of coverage.

Score_breakdown:

Alignment: 0.4

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.7

Reason: The score is 0.70 because the summary includes extra information not present in the original text, such as the ability for anyone to edit a wiki, the concept of a constantly updated platform, and the mention of building knowledge bases and documentation. Additionally, the summary fails to address whether most wikis include features like access rights management, which the original text can answer.

Score_breakdown:

Alignment: 0.7

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.8571428571428571

Reason: The score is 0.86 because the summary accurately reflects the original text without contradictions, although it includes additional details about robust access control and user management not specified in the original text. Overall, the summary is well-aligned with the original content.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.llama3_1_402b

File: summ_001_result.json

Score: 0.5555555555555556

Reason: The score is 0.56 because the summary contains contradictions regarding panel visibility and user control over panel width, includes extra information about the use of panels and wikis not found in the original text, and fails to answer a question about exporting pages in PDF format that the original text can address.

Score_breakdown:

Alignment: 0.5555555555555556

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains significant contradictions and adds multiple pieces of extra information not present in the original text. Additionally, it fails to answer specific questions that the original text can address, indicating a lack of fidelity and completeness in the summarization.

Score_breakdown:

Alignment: 0.5882352941176471

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.3333333333333333

Reason: The score is 0.33 because the summary contains a contradiction regarding the navigation method to the wiki administration page and includes numerous pieces of extra information not present in the original text, such as setting permissions for users and groups, priority order for access, and specific actions for guests, among others.

Score_breakdown:

Alignment: 0.3333333333333333

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.5714285714285714

Reason: The score is 0.57 because the summary contains multiple inaccuracies and omissions compared to the original text. It includes contradictions regarding the upgrade process and additional details not present in the original text. Furthermore, it fails to address specific questions that the original text can answer, indicating a lack of comprehensiveness and precision.

Score_breakdown:

Alignment: 0.5714285714285714

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.5

Reason: The score is 0.50 because the summary contains incorrect information about the Watchlist feature, suggesting it is a default RSS feed and used for creating customized feeds, which contradicts the original text. Additionally, the summary fails to address whether RSS feeds can be received for blog posts if the Blog application is installed, which the original text can answer.

Score_breakdown:

Alignment: 0.5

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.5

Reason: The score is 0.50 because the summary includes extra information not present in the original text, such as the division into sections for different roles, specific user actions like creating and editing a page, and a conclusion directing to the next step. Additionally, the summary fails to answer questions about logging in requirements and developers' capabilities, which the original text can address.

Score_breakdown:

Alignment: 0.5

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.8

Reason: The score is 0.80 because the summary is mostly accurate and concise, with no contradictions or unnecessary information added. However, it misses answering a specific question about whether all content added to XWiki is saved over time, indicating a slight gap in coverage.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains multiple inaccuracies and omissions. It contradicts the original text by mentioning a 'Registration' rights configuration and configurations like 'Editable/Open Wiki' that are not present in the original. Additionally, it introduces extra information about the next steps in the XWiki user guide that is not found in the original text. Furthermore, the summary fails to address a question about overriding global rights at the page level for unregistered users, which the original text can answer.

Score_breakdown:

Alignment: 0.4

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.6

Reason: The score is 0.60 because the summary includes extra information not present in the original text and fails to answer some questions that the original text can address. However, there are no contradictions, which is a positive aspect of the summary.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.5

Reason: The score is 0.50 because the summary includes extra information not present in the original text, such as specific features of XWiki and its suitability for education and a wide range of use cases, which were not mentioned in the original text.

Score_breakdown:

Alignment: 0.5

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.llama3_1_8b_Q4

File: summ_001_result.json

Score: 0.3333333333333333

Reason: The score is 0.33 because the summary contains significant contradictions regarding the layout of the page areas, includes extra information not present in the original text, and fails to answer a question that the original text can address. These issues indicate a lack of accuracy and completeness in the summarization.

Score_breakdown:

Alignment: 0.3333333333333333

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains significant contradictions regarding the priority and checking order of permissions, includes vague information not present in the original text, and fails to address specific questions that the original text can answer. These issues indicate a lack of accuracy and completeness in the summarization.

Score_breakdown:

Alignment: 0.75

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.3333333333333333

Reason: The score is 0.33 because the summary contains significant contradictions with the original text and includes a substantial amount of extra information not present in the original text. These issues indicate a lack of accuracy and relevance in the summary compared to the original content.

Score_breakdown:

Alignment: 0.3333333333333333

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.2

Reason: The score is 0.20 because the summary contains significant contradictions and adds a substantial amount of extra information not present in the original text. Additionally, it fails to address several questions that the original text can answer, indicating a poor alignment with the source material.

Score_breakdown:

Alignment: 0.2727272727272727

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.875

Reason: The score is 0.88 because the summary closely aligns with the original text, with only minor discrepancies such as mentioning specific browser extensions like Firefox, which were not specified in the original text. Overall, the summary is accurate and informative.

Score_breakdown:

Alignment: 0.875

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.4

Reason: The score is 0.40 because the summary includes extra information not present in the original text, such as links to guides and resources, specific guides on creating and editing pages, and invitations to proceed with certain steps. Additionally, the summary fails to answer questions that the original text can address, such as the necessity of logging in for certain actions, whether XWiki is a second-generation wiki, and the availability of additional applications on the Extensions wiki.

Score_breakdown:

Alignment: 0.42857142857142855

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.75

Reason: The score is 0.75 because the summary contains a contradiction regarding the location of the 'History' menu item and includes extra information about a 'Version Control page on XWiki.org' not present in the original text. Additionally, the summary fails to address whether all content added to XWiki is saved over time, which the original text can answer. Despite these issues, the summary is still relatively accurate and informative.

Score_breakdown:

Alignment: 0.75

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.8

Reason: The score is 0.80 because the summary includes extra information not present in the original text, such as screenshots or examples of configurations, and it fails to answer a question about specific configurations that the original text can address. However, there are no contradictions, indicating a generally accurate representation of the original content.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.8

Reason: The score is 0.80 because the summary is mostly accurate and concise, with no contradictions or unnecessary extra information. However, it lacks the ability to answer certain questions that the original text can, indicating room for improvement in comprehensiveness.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary includes extra information about features like blogs, task managers, community hubs, and education solutions that are not mentioned in the original text. However, there are no contradictions, which is a positive aspect of the summary.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.mistral-nemo_12b_Q4

File: summ_001_result.json

Score: 0.2

Reason: The score is 0.20 because the summary includes several inaccuracies and additional information not present in the original text. It incorrectly claims features like panel customization per space and multiple wiki creation, which are not mentioned in the original text. Furthermore, it introduces actions and functionalities that are not covered in the original text, leading to a significant deviation from the source material.

Score_breakdown:

Alignment: 0.2

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains multiple inaccuracies, including contradictions about editing rights, script execution, and administration rights, as well as extra information not present in the original text. Additionally, the summary fails to answer several questions that the original text can address, indicating a lack of coverage and accuracy.

Score_breakdown:

Alignment: 0.6

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.4

Reason: The score is 0.40 because the summary includes a significant amount of extra information not present in the original text, which affects its accuracy and relevance.

Score_breakdown:

Alignment: 0.4

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.14285714285714285

Reason: The score is 0.14 because the summary contains multiple contradictions with the original text, such as the incorrect use of the Distribution Wizard for upgrading flavors and the misrepresentation of recommended methods for upgrading and downgrading. Additionally, the summary includes extra information not found in the original text, like specific methods for upgrading and details about Solr issues, which further detracts from its accuracy. Furthermore, the summary fails to address several questions that the original text can answer, indicating a lack of comprehensiveness.

Score_breakdown:

Alignment: 0.14285714285714285

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.8888888888888888

Reason: The score is 0.89 because the summary includes extra information not present in the original text, specifically mentioning the Main.WebRss page for RSS feeds, which slightly affects the accuracy. However, there are no contradictions, indicating a generally high-quality summary.

Score_breakdown:

Alignment: 0.8888888888888888

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.6

Reason: The score is 0.60 because the summary includes incorrect information about the resources available to developers, adding the Extensions Wiki and XWiki Rendering Wiki, which are not mentioned in the original text. Additionally, the summary introduces extra information about users learning to create, edit, and view history of pages, which is not in the original text. Furthermore, the summary fails to address questions about starting a new wiki and whether XWiki is a second generation wiki, which the original text can answer.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains a contradiction regarding how users access the 'History' menu, includes extra information about XWiki.org's Version Control page not present in the original text, and fails to answer questions about content saving and the location of the 'History' tab that the original text can address.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.5555555555555556

Reason: The score is 0.56 because the summary contains multiple contradictions regarding the configurations described in the original text. It incorrectly combines various configurations such as Editable, Viewable, Hidden, and Protected with an Open configuration, which is not mentioned in the original text. This significantly impacts the accuracy and reliability of the summary.

Score_breakdown:

Alignment: 0.5555555555555556

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 1.0

Reason: The score is 1.00 because the summary perfectly aligns with the original text, with no contradictions or extra information present. This indicates a high-quality summarization task.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.8571428571428571

Reason: The score is 0.86 because the summary contains a minor contradiction regarding the location of built-in applications for XWiki. However, there is no extra information or unanswered questions, indicating that the summary is otherwise accurate and well-aligned with the original text.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.mistral2_large

File: summ_001_result.json

Score: 0.5454545454545454

Reason: The score is 0.55 because the summary contains contradictions regarding the creation of subwikis and the display of panels, includes extra information not present in the original text about actions available on a Page in XWiki and starting with a single wiki, and fails to answer questions about exporting pages and choosing panels, indicating a moderate level of fidelity to the original content.

Score_breakdown:

Alignment: 0.5454545454545454

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.7

Reason: The score is 0.70 because the summary contains contradictions regarding the availability of Programming Right and Script Right compared to the original text, and it includes extra information about a table in the Security Module documentation that is not mentioned in the original text.

Score_breakdown:

Alignment: 0.7

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.5333333333333333

Reason: The score is 0.53 because the summary contains significant contradictions and includes extra information not present in the original text. The contradiction regarding who can perform actions in an open wiki and the numerous pieces of extra information indicate a lack of alignment with the original content.

Score_breakdown:

Alignment: 0.5333333333333333

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.38461538461538464

Reason: The score is 0.38 because the summary contains several contradictions with the original text, such as the purpose of the Distribution Wizard and the method for handling Solr initialization. Additionally, it includes extra information not present in the original text, like upgrading from specific packages and steps for upgrading the WAR file. There is also a question that the original text can answer but not the summary, indicating a lack of comprehensiveness.

Score_breakdown:

Alignment: 0.38461538461538464

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 1.0

Reason: The score is 1.00 because the summary perfectly aligns with the original text, with no contradictions or extraneous information present. Excellent summarization quality.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.5

Reason: The score is 0.50 because the summary includes extra information not present in the original text, such as unspecified actions requiring login and directing users to create a page, which affects the accuracy and completeness of the summary.

Score_breakdown:

Alignment: 0.5

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.625

Reason: The score is 0.62 because the summary contains contradictions regarding the location and function of the 'History' tab and menu item, includes extra details not present in the original text such as a 'Version Control page on XWiki.org' and steps about changing the logo and panels, and fails to answer a question about content saving in XWiki that the original text can address.

Score_breakdown:

Alignment: 0.625

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.5555555555555556

Reason: The score is 0.56 because the summary contains several contradictions regarding user rights and registration settings that are not aligned with the original text, and it includes extra information not present in the original text. These issues indicate a moderate level of inaccuracy and incompleteness in the summary.

Score_breakdown:

Alignment: 0.5555555555555556

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.8

Reason: The score is 0.80 because the summary is mostly accurate and concise, with no contradictions or unnecessary information. However, it misses answering a specific question that the original text could address, indicating a slight gap in completeness.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.8

Reason: The score is 0.80 because the summary includes extra information about the types of new applications, such as blogs or task managers, that can be created in XWiki, which is not mentioned in

the original text. However, there are no contradictions, indicating a generally accurate summary.

Score_breakdown:

Alignment: 0.8

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.mixtral-8x22b

File: summ_001_result.json

Score: 0.4444444444444444

Reason: The score is 0.44 because the summary includes a significant amount of extra information not present in the original text, such as details about printing pages, creating child pages, and managing access rights. Additionally, the summary fails to address a question that the original text can answer, indicating a lack of coverage and accuracy.

Score_breakdown:

Alignment: 0.4444444444444444

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.7272727272727273

Reason: The score is 0.73 because the summary contains a contradiction regarding the similarity of 'Create Wikis Right' to programming rights, which is not mentioned in the original text. Additionally, the summary includes extra information about XWiki and a tabular view that is not present in the original text. Furthermore, the summary fails to answer a question about the default status of the comment right, which the original text can address.

Score_breakdown:

Alignment: 0.7272727272727273

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.5714285714285714

Reason: The score is 0.57 because the summary includes extra information not present in the original text, such as details about priority order for permissions, public or private configuration options, and specific documentation for setting sub-wiki access rights. Additionally, the summary fails to answer a question that the original text can, indicating a lack of completeness and accuracy.

Score_breakdown:

Alignment: 0.5714285714285714

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.5384615384615384

Reason: The score is 0.54 because the summary contains multiple contradictions and extra information not present in the original text, such as incorrect methods for upgrading and additional details about package updates and configuration backups. Additionally, the summary fails to address specific questions that the original text can answer, indicating a lack of alignment and completeness.

Score_breakdown:

Alignment: 0.5384615384615384

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.7

Reason: The score is 0.70 because the summary includes extra information not found in the original text, such as the Notifications Application, Atom, RDF, and customizing RSS content. However, there are no contradictions, which indicates a generally accurate summarization with some additional details.

Score_breakdown:

Alignment: 0.7

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.7272727272727273

Reason: The score is 0.73 because the summary includes extra information not present in the original text, such as user capabilities like creating a page, editing a page, and viewing page history, as well as the ability to create complex web applications. Additionally, it omits a detail about a link to the XWiki Development Zone. However, there are no contradictions, indicating a generally accurate representation of the original content.

Score_breakdown:

Alignment: 0.7272727272727273

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary includes extra information not present in the original text, such as using the browser's back button, a 'Version Control page on XWiki.org', and a step about changing the logo and panels. Additionally, there is a minor contradiction regarding the location of the 'Compare selected versions' button.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.75

Reason: The score is 0.75 because the summary includes extra information not present in the original text, such as leaving the Users Rights screen blank and a linked page on XWiki.org with more details and screenshots. However, there are no contradictions, indicating a generally accurate summary with some additional details.

Score_breakdown:

Alignment: 0.75

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.9

Reason: The score is 0.90 because the summary includes extra information not present in the original text, such as mentioning XWiki or reference pages. However, there are no contradictions, indicating a generally accurate summary with minor additions.

Score_breakdown:

Alignment: 0.9

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.8333333333333334

Reason: The score is 0.83 because the summary contains a contradiction regarding the location of built-in applications, which affects its accuracy. However, there is no extra information or unanswered questions, indicating a generally high-quality summary.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.phi3_medium-128k_14b_Q4

File: summ_001_result.json

Score: 0.5

Reason: The score is 0.50 because the summary contains contradictions regarding the structure and default settings of XWiki, such as the starting point of wikis and the default panel arrangement. Additionally, it includes extra information about unique panels per space that is not present in the original text. Furthermore, the summary fails to address specific questions about the functionality of panels and exporting options available in XWiki, which the original text can answer.

Score_breakdown:

Alignment: 0.5

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.23076923076923078

Reason: The score is 0.23 because the summary contains significant contradictions and introduces a substantial amount of extra information not present in the original text, indicating a low level of accuracy and relevance.

Score_breakdown:

Alignment: 0.23076923076923078

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.3

Reason: The score is 0.30 because the summary contains significant contradictions with the original text, such as incorrect details about permissions and admin options. Additionally, it includes extra information not present in the original text, like the comprehensive nature of the guide and specific details on rights management. Furthermore, the summary fails to address questions that the original text can answer, such as the privileges of the wiki owner and superadmin account.

Score_breakdown:

Alignment: 0.3

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.2

Reason: The score is 0.20 because the summary contains multiple contradictions with the original text, such as incorrect methods and tips for upgrades and downgrades. Additionally, it includes extra information not present in the original text and fails to answer specific questions that the original text can address. These issues significantly reduce the accuracy and usefulness of the summary.

Score_breakdown:

Alignment: 0.2

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.5

Reason: The score is 0.50 because the summary includes a contradiction about setting up alerts on the Main.WebRss page and adds extra information about XWiki documentation on RSS feeds that is not present in the original text.

Score_breakdown:

Alignment: 0.5

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.4

Reason: The score is 0.40 because the summary includes extra information not present in the original text, such as step-by-step instructions, specific actions requiring login, and details about developer features and tools. This additional content suggests a deviation from the original text, impacting the summary's accuracy.

Score_breakdown:

Alignment: 0.4

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.8

Reason: The score is 0.80 because the summary includes extra information not present in the original text and fails to answer a question that the original text can address. However, there are no contradictions, indicating a generally accurate representation of the original content.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.5

Reason: The score is 0.50 because the summary contains a contradiction regarding the conditions under which global rights can be overridden, includes extra information not present in the original text, and fails to address specific questions that the original text can answer.

Score_breakdown:

Alignment: 0.5

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.6

Reason: The score is 0.60 because the summary includes extra information not present in the original text, such as specific mentions of XWiki and its features, which were not part of the original content. Additionally, the summary fails to address questions about general wiki features like access rights management and notifications, which the original text could answer.

Score_breakdown:

Alignment: 0.6

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.5

Reason: The score is 0.50 because the summary includes contradictions regarding the location of application details and adds extra information about specific applications not mentioned in the original

text.

Score_breakdown:

Alignment: 0.5

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.phi3_mini-128k_4b_Q4

File: summ_001_result.json

Score: 0.375

Reason: The score is 0.38 because the summary contains several inaccuracies and contradictions with the original text, such as incorrect placement of title and author information, misrepresentation of panel placement, and misunderstanding of what constitutes a wiki. Additionally, the summary includes extra information not present in the original text, such as adjusting column sizes and selecting specific panels. Furthermore, the summary fails to address questions that the original text can answer, such as the use of panels for lateral menus in XWiki.

Score_breakdown:

Alignment: 0.375

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.36363636363636365

Reason: The score is 0.36 because the summary contains multiple contradictions with the original text, such as incorrect default statuses for various rights and misrepresentations of access requirements. Additionally, there is extra information in the summary not present in the original text, and it fails to answer a question that the original text can address.

Score_breakdown:

Alignment: 0.36363636363636365

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains significant contradictions and includes extra information not present in the original text. It inaccurately describes the hierarchy of permissions and introduces several elements not mentioned in the original text, such as guest user permissions and sub-wiki access rights. Additionally, it fails to answer questions that the original text can address, such as the precedence of page-level permissions over wiki-wide permissions.

Score_breakdown:

Alignment: 0.4

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.0

Reason: The score is 0.00 because the summary contains multiple contradictions and extra information not present in the original text. It inaccurately describes the use of the Distribution Wizard, downgrading processes, and troubleshooting tips for Solr lock-ups. Additionally, it includes details about maintaining data integrity, specific upgrade paths, and alternative approaches that are not mentioned in the original text. Furthermore, the summary fails to address questions about migration paths and the necessity of checking release notes, which the original text can answer.

Score_breakdown:

Alignment: 0.0

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.25

Reason: The score is 0.25 because the summary includes incorrect information about the RSS notification button being provided by default, adds extra details about extensions and guides not mentioned in the original text, and fails to address a question about RSS feed features related to the Watchlist.

Score_breakdown:

Alignment: 0.25

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.42857142857142855

Reason: The score is 0.43 because the summary contains contradictions regarding the content of the Developer's Guide, includes extra instructions and details not present in the original text, and fails to answer questions related to specific documentation resources that the original text can address.

Score_breakdown:

Alignment: 0.42857142857142855

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.5

Reason: The score is 0.50 because the summary contains incorrect information about the location of the 'History' tab and adds extra details about a user guide, URL, and tutorial that are not present in the original text.

Score_breakdown:

Alignment: 0.5

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.0

Reason: The score is 0.00 because the summary contains multiple contradictions with the original text, such as incorrect claims about user rights, registration settings, and permissions. Additionally, the summary includes extra information not present in the original text, like references to an Admin Guide and detailed instructions. Furthermore, the summary fails to answer questions that the original text can, such as whether XWiki allows admins to create new user accounts.

Score_breakdown:

Alignment: 0.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.5

Reason: The score is 0.50 because the summary contains contradictions regarding how changes to pages are handled, includes extra information not present in the original text, and fails to answer specific questions that the original text can address. These issues significantly impact the accuracy and completeness of the summary.

Score_breakdown:

Alignment: 0.5

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.8571428571428571

Reason: The score is 0.86 because the summary includes extra information about customized functionalities like blogs or product sheet managers that are not mentioned in the original text. However, there are no contradictions, indicating a generally accurate summary.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.qwen2_7b_Q4

File: summ_001_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains significant contradictions and extra information not found in the original text. It incorrectly describes the contents of the footer and includes additional details about XWiki that are not mentioned in the original text. Furthermore, the summary fails to answer questions that the original text can address, indicating a lack of coverage and accuracy.

Score_breakdown:

Alignment: 0.4444444444444444

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.7

Reason: The score is 0.70 because the summary contains contradictions regarding the scope and default status of certain rights, includes extra information about priority order not present in the original text, and fails to answer a question about the register right for a specific pseudo-user. These issues indicate a moderate level of accuracy and completeness in the summarization.

Score_breakdown:

Alignment: 0.7

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.14285714285714285

Reason: The score is 0.14 because the summary contains significant contradictions with the original text, such as the incorrect hierarchy of permissions and access restrictions. Additionally, the summary includes numerous pieces of extra information not found in the original text, and it fails to answer questions that the original text can address. These issues indicate a poor alignment between the summary and the original content.

Score_breakdown:

Alignment: 0.14285714285714285

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.16666666666666666

Reason: The score is 0.17 because the summary contains multiple inaccuracies and omissions. It includes contradictory information about the downgrading process, extra details not present in the original text, and fails to address several questions that the original text can answer.

Score_breakdown:

Alignment: 0.16666666666666666

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.9090909090909091

Reason: The score is 0.91 because the summary is highly accurate with no contradictions and only includes minor extra information about the flexibility of RSS feeds, which does not significantly detract from the overall quality.

Score_breakdown:

Alignment: 0.9090909090909091

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.42857142857142855

Reason: The score is 0.43 because the summary contains contradictions regarding the Developer's Guide and incorrectly suggests developers use pre-installed applications. Additionally, it includes extra information about creating and editing pages, page history, and suggestions for creating a page, which are not mentioned in the original text. Furthermore, the summary fails to answer a question about a dedicated wiki for XWiki Rendering documentation that the original text can address.

Score_breakdown:

Alignment: 0.42857142857142855

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary includes some incorrect information about the location of the 'More Actions' button and adds details not found in the original text, such as the 'Version Control page on XWiki.org' and changing the logo and panels in XWiki. Additionally, the summary fails to address whether all contents added to the wiki are saved over time, which the original text can answer.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.7777777777777778

Reason: The score is 0.78 because the summary includes extra information not present in the original text, such as references to images and a specific step in a process. However, there are no contradictions, indicating a generally accurate representation of the original content.

Score_breakdown:

Alignment: 0.7777777777777778

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 1.0

Reason: The score is 1.00 because the summary perfectly aligns with the original text, with no contradictions or extra information present.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.5714285714285714

Reason: The score is 0.57 because the summary contains a contradiction regarding where the list of built-in applications is detailed. Additionally, the summary includes extra information about contexts and adaptability that are not mentioned in the original text.

Score_breakdown:

Alignment: 0.5714285714285714

Coverage: 1.0

Input_language: en

Summary_language: en

Text_generation Results

Model: AI.Models.GPT4o

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output contains all the key information accurately but is not in the JSON format as the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output is relevant and coherent, addressing the Input prompt by using all keywords in a logical sentence. It maintains fluency and quality similar to the Expected Output, with only a slight variation in the verb used, which does not impact overall quality.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively simplifies the concept of photosynthesis for a younger audience, similar to the Expected Output. It maintains the key theme of converting sunlight into energy and relates it to children's experiences. The language is engaging and age-appropriate, with logical flow and grammatical correctness.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, directly addressing the features and target audience. It maintains coherence with a logical flow similar to the Expected Output, covering all key points like noise-cancelling, battery life, and comfort. The fluency is excellent, with smooth and readable text, closely matching the language quality of the Expected Output.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_005_result.json

Score: 0.9

Reason: The Actual Output is a creative story about a space adventure, directly addressing the Input. It aligns with the Expected Output by featuring a group of astronauts exploring a new planet, although it expands on the theme with additional elements like interstellar travel and alien encounters. The narrative is coherent with well-connected ideas, and the language is grammatically accurate and fluent.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent poem about the beauty of nature, aligning perfectly with the Input and Expected Output by describing the wonders of the natural world through vivid imagery and themes.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent news article that directly addresses the Input prompt by detailing a recent scientific discovery in quantum computing. It aligns with the Expected Output by summarizing the breakthrough effectively, maintaining logical flow and clarity throughout.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output is a comprehensive travel guide for Kyoto, Japan, aligning perfectly with the Input prompt's requirement to create a travel guide for a popular tourist destination. It covers top attractions, accommodations, activities, and additional travel tips, meeting the Expected Output's criteria. The content is coherent, fluent, and well-structured, effectively addressing all evaluation steps.

Prompt_language: ro
Expected_answer_language: en
Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output accurately captures the event name, dates, location, and features, aligning with the Input. However, it combines 'venue' and 'address' into one line and uses different labels compared to the Expected Output, slightly affecting coherence.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the Input for a technical audience by using precise terminology such as 'physiological processes' and 'cognitive functions.' It maintains coherence and fluency similar to the Expected Output, though it slightly lacks the depth of detail found in the Expected Output's mention of 'subjective experiences.'

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

Model: AI.Models.GPT4o-mini

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output accurately extracts the key information from the Input but does not match the Expected Output format, which is a JSON object.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output is relevant to the Input prompt, coherent, and fluent. It closely matches the Expected Output, capturing the key elements and intent, with only a minor difference in wording.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience by using relatable language like 'magic trick' and 'food,' which aligns with the Input prompt. It maintains coherence and fluency with a logical flow and clear vocabulary. While it introduces the idea of energy for 'playing and having fun,' which is not in the Expected Output, it still conveys the core idea of energy conversion in a child-friendly manner.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, addressing all key features and the target audience. It maintains coherence and logical flow similar to the Expected Output, with slight variations

in phrasing. The fluency is excellent, with clear and grammatically correct sentences. It meets the Expected Output in all aspects but uses different language.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.8

Reason: The Actual Output is a creative space adventure story, aligning with the Input prompt. It is coherent and fluent, with a clear structure and logical progression. However, it diverges from the Expected Output by focusing on a galaxy-wide quest rather than solely exploring a new planet.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent poem that directly addresses the Input prompt by vividly describing the beauty of nature, aligning perfectly with the Expected Output's requirement of depicting the wonders of the natural world.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output is highly relevant to the Input prompt, providing a detailed news article on a recent scientific discovery. It is coherent, with a logical flow and connection of ideas, and fluent, with correct grammar and natural language. It meets the Expected Output by summarizing a groundbreaking scientific finding effectively.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, providing a comprehensive travel guide to Kyoto, a popular tourist destination. It covers attractions, transportation, cultural experiences, and tips, aligning well with the Expected Output. The content is coherent, fluent, and logically structured. However, it lacks specific mention of accommodations, which slightly deviates from the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output accurately extracts the event name, date, location, and features, addressing the Input. It aligns well with the Expected Output in terms of main ideas but uses a different format and lacks explicit separation of start and end dates.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output closely aligns with the Expected Output, effectively adapting the text for a technical audience by using terms like 'neurobiological organ' and 'cognitive processes'. It maintains coherence and fluency, with only minor differences in phrasing such as 'central substrate' instead of 'biological substrate'.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.claude3_5_sonnet

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output accurately extracts the key information from the Input and presents it clearly, maintaining relevance and coherence. However, it does not match the JSON format of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input prompt and aligns well with the Expected Output in terms of relevance. It maintains coherence and logical flow, similar to the Expected Output. The language is fluent and grammatically correct, matching the naturalness of the Expected Output. Minor differences in wording do not affect the overall quality.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience by using relatable language and examples, such as comparing plants to food factories and likening photosynthesis to eating lunch. It maintains relevance and coherence with the Input and Expected Output, though it adds extra analogies not present in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, addressing key themes like noise-cancelling, battery life, and comfort for tech-savvy professionals. It is coherent and fluent, with a logical flow and good grammar. While it slightly differs in wording from the Expected Output, it meets the key elements and quality criteria effectively.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.7

Reason: The Actual Output is relevant to the Input prompt as it is a creative space adventure story. However, it diverges from the Expected Output by focusing on a space anomaly and first contact rather than exploring a new planet. The story is coherent and fluent, but the discrepancy in the setting affects alignment with the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent poem that directly addresses the Input prompt by celebrating the beauty of nature, aligning perfectly with the Expected Output of describing the wonders of the natural world.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input by presenting a news article about a scientific discovery, aligning with the Expected Output of summarizing a groundbreaking finding. It is coherent, logically structured, and fluent, with clear sentence structure and grammar. The fictional note slightly detracts from full realism but does not significantly impact the article's quality.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output provides a comprehensive travel guide for Paris, addressing key points such as attractions, accommodations, and activities, aligning well with the Expected Output. It is detailed and coherent, with logical flow and relevant information.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output addresses the Input prompt by extracting relevant event details and aligns with the Expected Output in terms of topic relevance. It is coherent and logically organized, with a slight variation in format compared to the Expected Output. The fluency is maintained with correct grammar and natural language use. However, it includes additional information not present in the Expected Output, such as 'Type' and 'Features', and doesn't separate the start and end dates.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the input for a technical audience by using precise terminology like 'central processing unit,' 'afferent sensory inputs,' and 'neurochemical signaling pathways,' aligning well with the Expected Output's focus on physiological processes and cognitive functions. It maintains coherence and fluency with complex sentence structures and advanced vocabulary.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.command-r_35B_Q4

File: text_gen_001_result.json

Score: 0.9

Reason: The Actual Output accurately extracts and presents the key information from the Input. It is relevant, coherent, and fluent. The only discrepancy is the use of colons instead of JSON format in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output is relevant to the Input prompt, transforming the keywords into a coherent sentence. It maintains logical flow and consistency with the Expected Output, though it uses 'frolicked' instead of 'enjoyed exploring'. The sentence structure and readability are fluent and match the quality of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively simplifies the concept of photosynthesis for a younger audience, using engaging language and relatable analogies. It maintains coherence and logical flow, aligning well

with the Expected Output. However, it slightly embellishes with phrases like 'superpower,' which, while engaging, diverges a bit from the straightforward nature of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output is relevant to the Input, addressing all features like noise-cancelling, 20-hour battery life, and comfortable fit, targeted at tech-savvy professionals. It is coherent with a logical flow and consistent ideas, similar to the Expected Output. The style is fluent and natural, matching the Expected Output's readability.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.8

Reason: The Actual Output is relevant to the Input prompt as it tells a creative space adventure story. It is coherent and fluent, with a logical flow and good grammar. However, it focuses on a single astronaut, Luna, rather than a group as mentioned in the Expected Output, which slightly misaligns with the expected narrative.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, addressing the beauty of nature as requested. It is coherent, with a logical flow and consistent ideas, aligning well with the Expected Output. Language and grammar are smooth and natural, closely matching the Expected Output's fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, providing a detailed article on a recent scientific discovery. It is coherent, with a logical flow and consistent ideas, matching the Expected Output's requirement for an informative article. Minor grammatical errors like 'serendetically' slightly affect fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output provides a comprehensive travel guide for Paris, directly addressing the Input prompt. It includes detailed sections on landmarks, restaurants, accommodations, and travel tips, aligning with the Expected Output's requirement for an informative guide. The language is fluent and coherent, maintaining a logical flow throughout the guide.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output addresses the main topic and includes most key details like event name, date, and location. However, it lacks separation of start and end dates and doesn't explicitly mention 'target_audience' as in the Expected Output. The output is coherent and fluent.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.8

Reason: The Actual Output addresses the main topics of the Input, such as brain functions and interpretation of external data, aligning well with the Expected Output. It maintains coherence and logical flow, though it simplifies some technical terms like 'physiological processes' and 'biological substrate'. The language is fluent and consistent with a technical audience.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.gemma2_9B_Q4

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output accurately extracts the key information from the Input and aligns with the Expected Output in terms of content. However, the format differs slightly as it uses bullet points instead of JSON structure.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output maintains the main topic and context of the Input, with all keywords included. It is coherent with a logical flow similar to the Expected Output, though the order of phrases differs slightly. The fluency is high, with correct grammar and syntax.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience by using relatable analogies, addressing the Input prompt. It maintains coherence and fluency, though it adds more detail than the Expected Output, which is more concise.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output aligns well with the Input prompt, covering all main features such as noise-cancelling, battery life, and comfort, targeting tech-savvy professionals. It maintains coherence with a logical flow similar to the Expected Output, though it uses a slightly different structure. The fluency is high, with clear and grammatically correct sentences, closely matching the Expected Output's readability.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.8

Reason: The Actual Output is a creative space adventure story, aligning well with the Input prompt. It is coherent and fluent, with a logical flow and engaging narrative. However, it focuses on an anomaly rather than directly exploring a new planet, which slightly diverges from the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is highly relevant to the Input, as it is a poem about the beauty of nature. It is coherent with a logical flow and consistent ideas, aligning well with the Expected Output of describing the wonders of the natural world. The language is fluent, with excellent grammar and style, matching the quality and readability expected.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output is highly relevant to the Input prompt, providing a detailed news article on a recent scientific discovery about Europa's ocean. It is coherent, with logically connected ideas, and fluent, with clear language and grammar. The key elements and intent of the Expected Output are met, as it summarizes a groundbreaking finding effectively.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output is a comprehensive travel guide for Florence, directly addressing the Input prompt. It includes detailed sections on attractions, accommodations, and activities, aligning with the Expected Output. The content is coherent, logically structured, and fluent, matching the quality expected.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output accurately extracts and presents the event details, maintaining relevance and coherence. It uses natural language and proper grammar. However, it differs slightly from the Expected Output in formatting and structure, such as combining start and end dates and including additional activities information.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the text for a technical audience by using terms like 'neural network' and 'synapses,' aligning well with the Expected Output. It maintains coherence and fluency, though it slightly diverges in phrasing compared to the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.llama3_1_402b

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output accurately extracts the key information from the Input, matching the Expected Output in content. However, the format differs slightly as it uses a list format instead of a JSON object.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.8

Reason: The Actual Output is relevant and coherent, addressing the main topic with logical flow. It is grammatically correct and natural. However, it introduces 'chased after butterflies' instead of 'enjoyed exploring,' which slightly deviates from the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience by simplifying the language and adding engaging elements, such as comparing it to a special power. It addresses the Input prompt well and maintains coherence with the Expected Output by explaining the process and its purpose. The fluency is enhanced with a friendly tone, making it suitable for ages 8-10.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, addressing all features and the target audience. It maintains coherence with a logical flow similar to the Expected Output, covering noise-cancelling, battery life, and comfort. The fluency is excellent, with proper grammar and syntax, matching the Expected Output's quality.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.8

Reason: The Actual Output is a creative space adventure story that aligns with the Input prompt. It is relevant and coherent, with a logical flow and consistent ideas. However, it diverges from the Expected Output by focusing on a cosmic exploration rather than a specific new planet, which slightly affects relevance. The language is fluent and engaging, matching the Expected Output's fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent poem that directly addresses the Input prompt by vividly describing the beauty of nature. It aligns well with the Expected Output by capturing the wonders of the natural world through imagery and thematic elements.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output is a well-structured news article about a recent scientific discovery, directly addressing the Input prompt. It provides coherent and detailed information about the discovery of Homo luzonensis, matching the Expected Output's requirement for an informative summary. The language is

fluent, with clear sentence structure and grammar, aligning with the quality expected.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output is highly relevant to the Input prompt as it provides a detailed travel guide for Santorini, a popular tourist destination. The content is coherent, well-organized, and covers key aspects such as attractions, accommodations, and activities, aligning perfectly with the Expected Output. Additionally, the language is fluent and grammatically correct.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output accurately extracts the event name, date, and location, aligning well with the Input. However, it lacks specific details such as separating the start and end dates, and omits the target audience, which are present in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the text for a technical audience, maintaining relevance to the Input by addressing the brain's complexity and functions. It is coherent and fluent, with logical flow and correct grammar. While it captures the key elements of the Expected Output, it slightly differs in phrasing, such as 'facilitating cognitive processes' instead of 'biological substrate for cognitive functions.'

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.llama3_1_8b_Q4

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output contains all the key information from the Input and aligns with the main ideas. However, it does not match the Expected Output format, lacking JSON structure, which affects coherence.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output is coherent and fluent, closely aligning with the Expected Output. Both mention a sunny day, a playful cat, and a garden. The Actual Output is slightly more detailed but maintains relevance and logical flow.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience by using simple language and relatable analogies, addressing the key theme of how plants use sunlight for energy. It maintains coherence and fluency, though it adds extra details about oxygen that are not in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output addresses the main topics of noise-cancelling, battery life, and comfort for tech-savvy professionals. It is coherent and fluent, with logical flow and correct grammar. The Expected Output aligns well with the Input and is consistent with the Actual Output, though the Actual Output uses slightly different phrasing.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.8

Reason: The Actual Output is a creative space adventure story, addressing the main topic of the Input. It includes exploration and adventure themes consistent with the Expected Output. However, it focuses on a wormhole and ancient temple rather than solely exploring a new planet. The story is coherent, grammatically correct, and fluent.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent poem that directly addresses the Input prompt by vividly describing the beauty of nature, aligning well with the Expected Output's theme of the wonders of the natural world.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the topic of a recent scientific discovery, aligning with the Input request. It matches the Expected Output by providing an informative summary of a groundbreaking finding in quantum physics. The article is coherent, with a logical flow of ideas, and is grammatically correct and readable.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output is a comprehensive travel guide for Paris, aligning perfectly with the Input prompt. It covers top attractions, accommodations, activities, and practical tips, matching the Expected Output's requirement for an informative guide on a popular tourist destination. The content is coherent, fluent, and relevant, addressing all evaluation steps effectively.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output correctly extracts the event name, dates, venue, and address, aligning with the Expected Output. However, it omits the target audience, which is a key theme in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.8

Reason: The Actual Output effectively uses technical terms and maintains relevance to the Input prompt. It introduces terms like 'cerebral cortex' and 'afferent inputs' which align with a technical audience. However, it diverges slightly from the Expected Output by emphasizing 'spiritual experiences' instead of 'subjective experiences' and using 'central processing unit (CPU)' metaphorically, which may not align perfectly with the Expected Output's focus on biological substrates.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.mistral-nemo_12b_Q4

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output addresses the main topic by extracting the required information, but it does not match the JSON format of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.7

Reason: The Actual Output addresses the Input prompt and includes all keywords, aligning with the Expected Output in relevance. However, it lacks coherence and fluency compared to the Expected Output, as it does not emphasize 'playful' and changes the sentence structure. The tone and style are consistent.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience by using relatable analogies and simple language. It maintains coherence and fluency throughout. However, it includes additional details like chloroplasts and oxygen release not present in the Expected Output, which focuses solely on energy conversion.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output addresses the Input prompt and aligns well with the Expected Output by highlighting noise-cancelling, 20-hour battery life, and comfort for professionals. It is coherent and fluent, though it lacks the explicit mention of 'tech-savvy' and 'ergonomic fit' present in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.6

Reason: The Actual Output is a creative space adventure story, aligning with the Input prompt. However, it diverges from the Expected Output as it focuses on music and a nebula rather than exploring a new planet. The story is coherent and fluent, with a logical flow and high-quality language.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent poem that beautifully describes the wonders of the natural world, directly addressing the Input prompt and meeting the Expected Output's intent.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output is highly relevant to the Input as it provides a detailed news article about a recent scientific discovery. It is coherent, with logically connected ideas and smooth flow. The language is fluent, with correct grammar and appropriate vocabulary. The Actual Output aligns well with the Expected Output by summarizing a groundbreaking scientific finding effectively.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output provides a comprehensive travel guide for Paris, aligning perfectly with the Input prompt and Expected Output. It covers attractions, accommodations, activities, and travel tips, maintaining coherence and fluency throughout.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output accurately captures the event name, date, venue, and address from the Input, maintaining relevance and coherence. However, it omits the 'target_audience' detail present in the Expected Output, slightly affecting completeness.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.8

Reason: The Actual Output is relevant to the Input, addressing the complexity and functions of the brain. It maintains coherence with the Expected Output, though it uses slightly different terminology. The fluency is high, with a natural and readable style, but it lacks some of the technical specificity found in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.mistral2_large

File: text_gen_001_result.json

Score: 0.9

Reason: The Actual Output accurately extracts and presents the key information from the Input. However, it uses a different format than the Expected Output, which affects coherence.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output is relevant to the Input prompt, maintaining coherence and fluency. It uses similar keywords and structure as the Expected Output, with minor differences in verb tense and word choice.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience by using simple language and relatable analogies, maintaining relevance to the Input. It is coherent and logically structured, similar to the Expected Output, but adds more engaging elements. The fluency is high, with natural and readable language suitable for ages 8-10.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, addressing all product features and the target audience. It maintains coherence with the Expected Output, providing a logical flow and consistent details. The fluency is excellent, with clear sentence structure and grammar, closely matching the Expected Output's quality.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_005_result.json

Score: 0.8

Reason: The Actual Output aligns with the Input by presenting a space adventure, and it introduces a group exploring new worlds, similar to the Expected Output. However, the Expected Output specifically mentions astronauts exploring a new planet, while the Actual Output involves a broader cosmic journey with a space pirate conflict. The structure and fluency are coherent and grammatically accurate, but the thematic focus slightly diverges from the Expected Output.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent poem that directly addresses the Input prompt by vividly describing the beauty of nature, aligning well with the Expected Output of depicting the wonders of the natural world.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by providing a detailed news article about a recent scientific discovery, specifically a breakthrough in fusion energy. It aligns with the Expected Output by summarizing the finding in an informative manner. The article is coherent, logically structured, and maintains consistency throughout. Additionally, it is fluent, using grammatically correct and natural language, matching the Expected Output's fluency level.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output is a comprehensive travel guide for Bali, directly addressing the Input prompt and aligning with the Expected Output by highlighting top attractions, accommodations, and activities. It is coherent, fluent, and exceeds the quality criteria by providing detailed and relevant information.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.9

Reason: The Actual Output accurately extracts all event details from the Input, maintaining relevance and coherence. It includes event name, dates, venue, address, features, and audience, aligning closely with the Expected Output. The only minor discrepancy is the format of the dates and key naming, which slightly affects fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output closely follows the structure and content of the Expected Output, using technical language suitable for the audience. It maintains coherence and fluency, with only slight variations in phrasing such as 'manifestation of cognitive faculties' instead of 'serving as the biological substrate for cognitive functions.'

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.mixtral-8x22b

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output correctly extracts and presents the key information from the Input, maintaining relevance. However, it does not match the Expected Output's JSON format, affecting coherence and fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output is relevant to the Input prompt, maintaining the same topic and requirements. It is coherent with a logical flow and consistent ideas, similar to the Expected Output. The only minor difference is the use of 'enjoys' instead of 'enjoyed', which slightly affects fluency but does not significantly impact readability.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience by using relatable analogies and simple language, aligning well with the Expected Output. It maintains coherence and fluency, though it introduces the idea of 'magic trick' which is not present in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, addressing all key features and the target audience. It is coherent with a logical flow and maintains fluency with correct grammar and vocabulary. It closely aligns with the Expected Output, though it adds slightly more promotional language.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent creative story about a space adventure, directly addressing the Input prompt. It follows a group of astronauts exploring a new planet, aligning well with the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent poem that vividly describes the beauty of nature, aligning perfectly with the Input prompt and the Expected Output of describing the wonders of the natural world.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output is highly relevant to the Input prompt, aligning with the topic of a recent scientific discovery. It is coherent, with a logical flow and consistency of ideas, matching the Expected Output's requirement for an informative article. The language is fluent, grammatically correct, and natural, meeting the quality expected.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input by providing a comprehensive travel guide for Paris, a popular tourist destination, aligning with the Expected Output by highlighting attractions, accommodations, and activities. It is coherent, logically structured, and maintains consistency with the Input and Expected Output. The language is fluent, grammatically correct, and uses natural expressions.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output accurately extracts event details such as event name, date, and location, and includes additional details about the event. However, it does not separate the start and end dates or specifically identify the target audience as in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output effectively addresses the Input by using technical terms like 'cerebrum' and 'neural network,' aligning well with the Expected Output's focus on complexity and function. It maintains coherence and fluency with logical flow and grammatical correctness, though it slightly deviates by emphasizing 'CPU' and 'locus of cognition' instead of 'biological substrate.'

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.phi3_medium-128k_14b_Q4

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output correctly extracts the key information from the Input and is relevant, but it does not match the JSON format of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output is relevant to the Input prompt and coherent, with ideas logically connected. It fluently describes the scene using all keywords. The difference from the Expected Output is minor, with 'frolicking in the vibrant garden' instead of 'exploring the garden'.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.8

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience, using engaging language and relatable analogies. It introduces the idea of plants making food with sunlight, similar to the Expected Output, but adds more detail and creativity. However, it slightly diverges from the Expected Output by including additional elements like 'solar-powered kitchen' and 'tasty vegg groves,' which, while imaginative, are not present in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output effectively addresses the topic by targeting tech-savvy professionals and highlighting noise-cancelling, 20-hour battery life, and comfort. It is coherent, fluent, and aligns well with the Expected Output, though it includes additional details about wireless connectivity and style.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input prompt, providing a detailed and creative story about a space adventure. It is coherent with a logical flow of events and maintains fluency with grammatically correct and natural language. It exceeds the Expected Output by offering a rich narrative

with character development and plot depth, while still meeting the core expectation of astronauts exploring a new planet.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent poem that vividly describes the beauty and wonders of nature, aligning perfectly with the Input prompt and Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input prompt, providing a detailed article on a scientific discovery about black holes, aligning with the Expected Output. It is coherent and fluent, with a logical flow and clear language. The only minor issue is the typo in the date, '2e45', which slightly affects the fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input prompt and Expected Output, providing a comprehensive travel guide to Paris with detailed sections on attractions, accommodations, and activities. It is coherent with a logical flow and consistent structure, covering essential topics expected in a travel guide. The fluency is mostly excellent, with minor grammatical errors like 'ann cups' instead of 'annual visits.'

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.3

Reason: The Actual Output partially addresses the Input by extracting some event details, but includes irrelevant content about flip flops. It lacks coherence and logical flow compared to the Expected Output, which clearly structures the event details. The fluency is disrupted by the unrelated content, affecting readability.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.8

Reason: The Actual Output effectively addresses the Input by using technical language and expanding on the brain's functions, aligning well with the Expected Output. It maintains coherence and fluency but includes more detail than necessary, slightly deviating from the concise nature of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.phi3_mini-128k_4b_Q4

File: text_gen_001_result.json

Score: 0.9

Reason: The Actual Output accurately extracts the key information from the Input and matches the Expected Output in content. However, the format differs slightly as it uses a list format instead of a JSON object.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.8

Reason: The Actual Output captures the essence of the Input and is relevant to the Expected Output. It is coherent and fluent, but it adds more detail than necessary, deviating slightly from the simplicity of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.8

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience by using relatable analogies and simpler language, addressing the main topic. It is coherent and fluently explains the process, though it is more detailed than the Expected Output. The main discrepancy is the added complexity which might slightly exceed the comprehension level of the target age group.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, addressing the product features and target audience effectively. It maintains coherence with a logical flow, aligning well with the Expected Output's structure. The language is fluent, with minor verbosity compared to the Expected Output, but it remains readable and grammatically correct.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.9

Reason: The Actual Output is a creative story about a space adventure, aligning well with the Input prompt. It is coherent and fluent, with a logical flow and grammatical correctness. It expands on the Expected Output by detailing the exploration of a new planet, Elysium VI, by astronauts, specifically focusing on Amelia Green's journey. The story exceeds the simplicity of the Expected Output by adding depth and complexity, but remains relevant and consistent with the theme of space exploration.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 0.9

Reason: The Actual Output is a coherent and fluent poem that beautifully captures the essence of nature, aligning well with the Input prompt and Expected Output. It uses vivid imagery and metaphor to describe the wonders of the natural world, meeting the intended goals effectively.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, providing a detailed news article about a scientific discovery, which aligns with the Expected Output. It maintains coherence and logical flow, detailing the discovery, its implications, and expert opinions. The fluency is strong, with appropriate sentence structure and vocabulary. However, it is slightly verbose, which could detract from conciseness.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 0.9

Reason: The Actual Output provides a detailed travel guide for Paris, addressing the Input request effectively. It maintains logical flow and consistency, covering attractions, accommodations, and activities, which aligns with the Expected Output. The language is fluent and free of grammatical errors. However, it includes promotional content which slightly deviates from a purely informative guide.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.7

Reason: The Actual Output captures the event name, dates, and location address accurately from the Input. However, it misses separating the start and end dates, specifying the venue name, and mentioning the target audience as seen in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.8

Reason: The Actual Output effectively adapts the Input for a technical audience by using precise terminology and detailed explanations, aligning well with the Expected Output's complexity. However, it is slightly verbose and less concise than the Expected Output, which affects coherence and fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.qwen2_7b_Q4

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output correctly extracts the key information from the Input and aligns with the Expected Output in terms of relevance and coherence, but it lacks the JSON format required by the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.8

Reason: The Actual Output is relevant to the Input, addressing the main topic of a cat in a garden on a sunny day. It is coherent, but lacks the adjective 'playful' from the Expected Output, affecting the alignment. The fluency is good, matching the Expected Output's quality.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively adapts the concept of photosynthesis for a younger audience by using simple language and engaging imagery, aligning well with the Input prompt. It maintains coherence and fluency, with logical connections and correct grammar. While it introduces creative elements like 'secret recipe' and 'superpower,' it still conveys the main idea of photosynthesis, similar to the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input prompt by describing the wireless Bluetooth headphones with noise-cancelling, 20-hour battery life, and comfortable fit, targeting tech-savvy professionals. It maintains coherence and fluency with a logical flow and smooth language. While it

includes additional details like touch controls and a charging case not mentioned in the Expected Output, it still meets the key elements and intent of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.9

Reason: The Actual Output is a creative story about a space adventure, aligning well with the Input prompt. It is coherent, with a logical flow from the mission's start to the discovery of ancient beings. The language is fluent and grammatically correct. Although it exceeds the Expected Output by including more details about the aliens and diplomacy, it still maintains relevance to the theme of exploring a new planet.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is a coherent and fluent poem that directly addresses the Input prompt by beautifully describing the wonders of the natural world, aligning perfectly with the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output is highly relevant to the Input, providing a detailed news article about a recent scientific discovery in neuroscience. It is coherent, with logically connected ideas and smooth, natural language. The key elements and intent of the Expected Output are met, as the article is informative and summarizes a groundbreaking scientific finding effectively.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 0.9

Reason: The Actual Output provides a comprehensive travel guide for Bali, addressing the Input prompt effectively with detailed sections on transportation, accommodation, attractions, and more. It maintains a logical flow and consistency in ideas, aligning well with the Expected Output of highlighting top attractions and activities. The fluency is high with clear sentence structure and grammar, matching the quality expected.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output accurately extracts event details such as name, dates, location, and activities, aligning well with the Input. However, it misses the specific start and end dates and the target audience, which are present in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, addressing the complexity and functions of the brain with technical details like neuron count. It maintains coherence with the Expected Output by discussing cognitive functions and consciousness, though it introduces additional philosophical elements. The fluency is consistent with the Expected Output, using sophisticated language appropriate for a technical audience.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en