

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 |

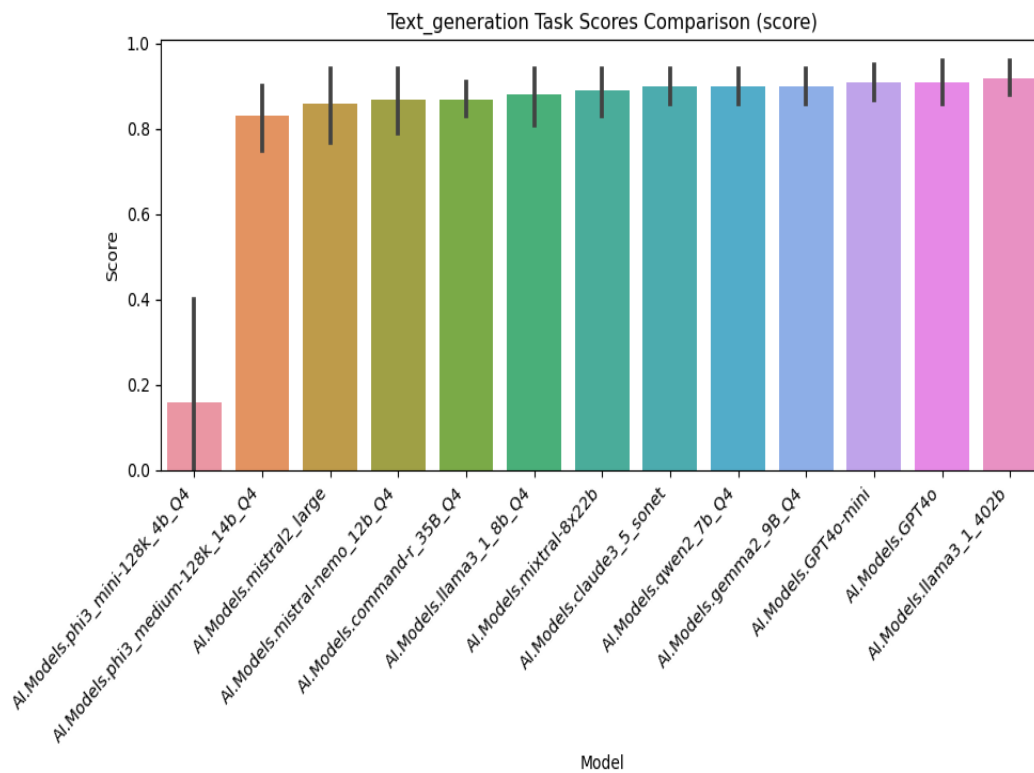
Setup Information

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

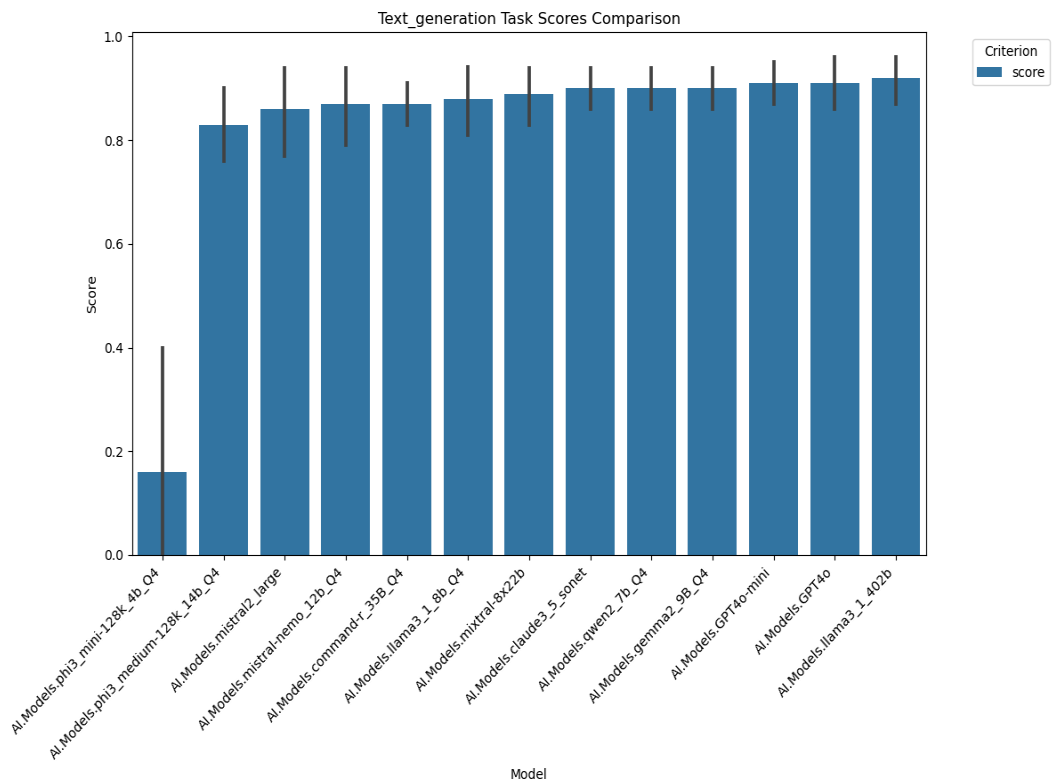
| | | | |
|---------------|--------------------------------------|------|-----|
| RAG-qa | AI.Models.qa_GPT4o | 0.8 | No |
| RAG-qa | AI.Models.qa_GPT4o-mini | 0.8 | No |
| RAG-qa | AI.Models.qa_claude3_5_sonnet | 0.8 | No |
| RAG-qa | AI.Models.qa_mistral2_large | 0.8 | No |
| RAG-qa | AI.Models.qa_llama3_1_402b | 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 |

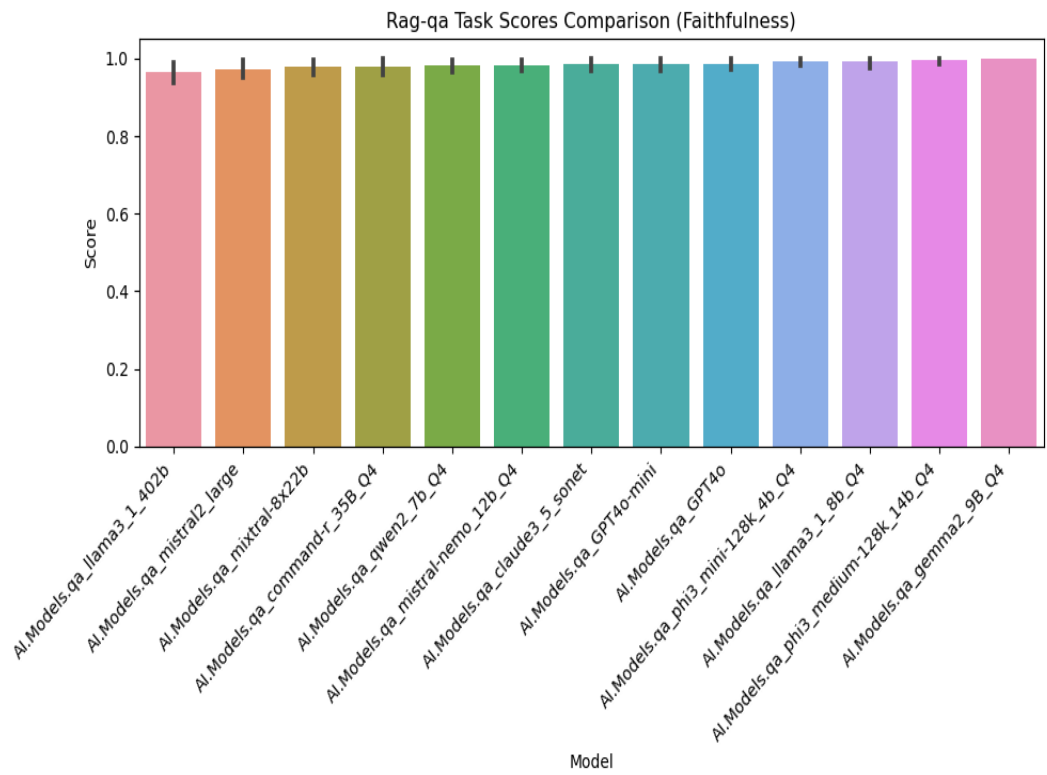
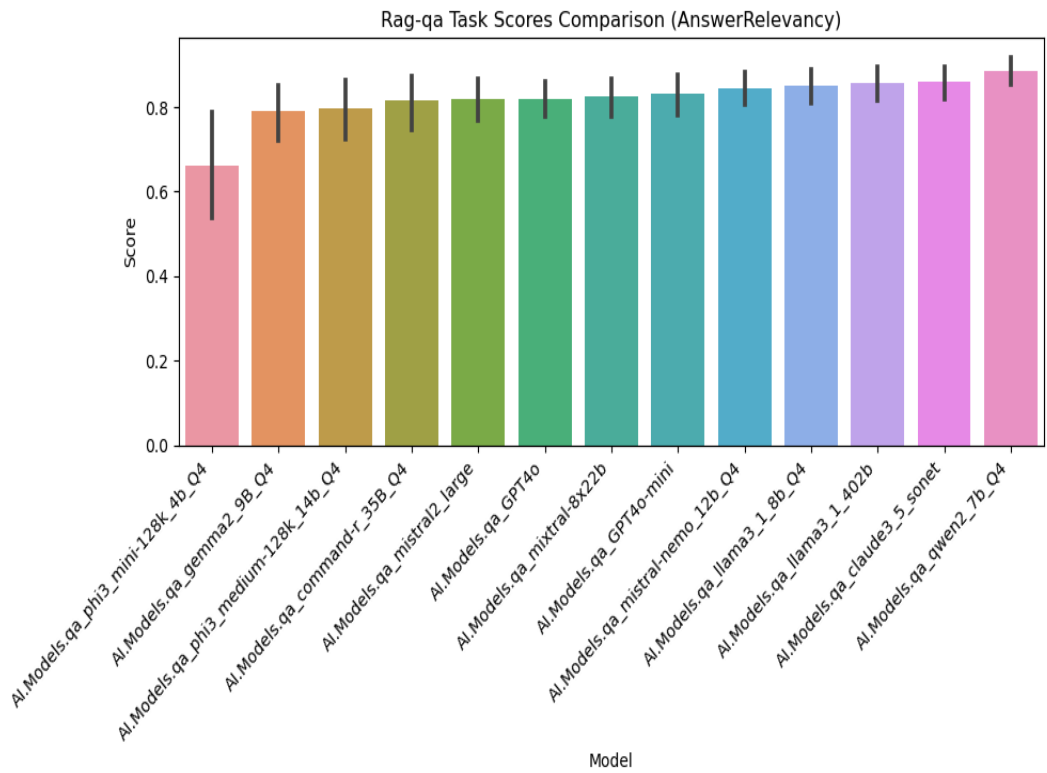
Text_generation Task Results

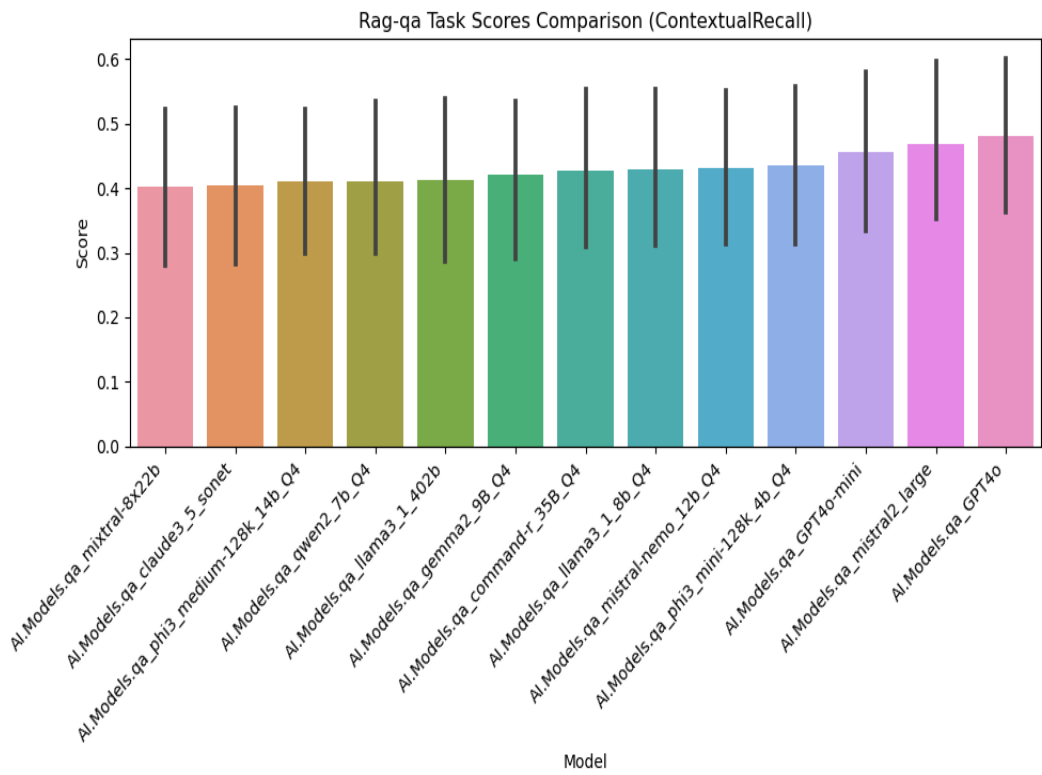
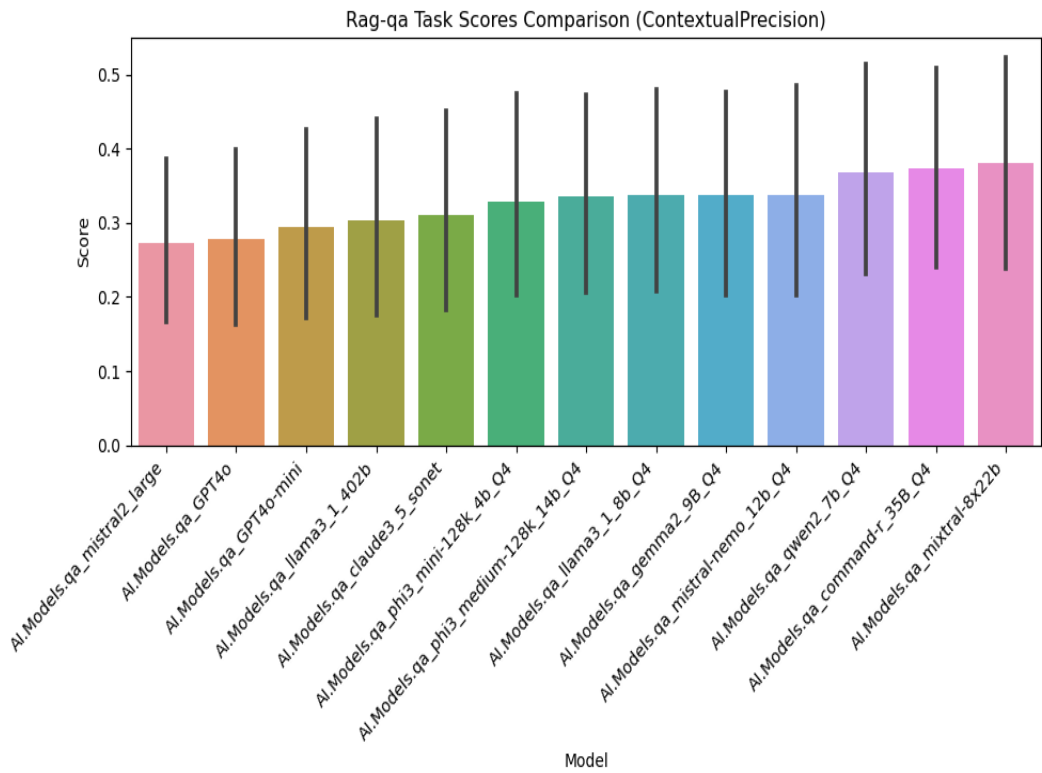


Grouped Results

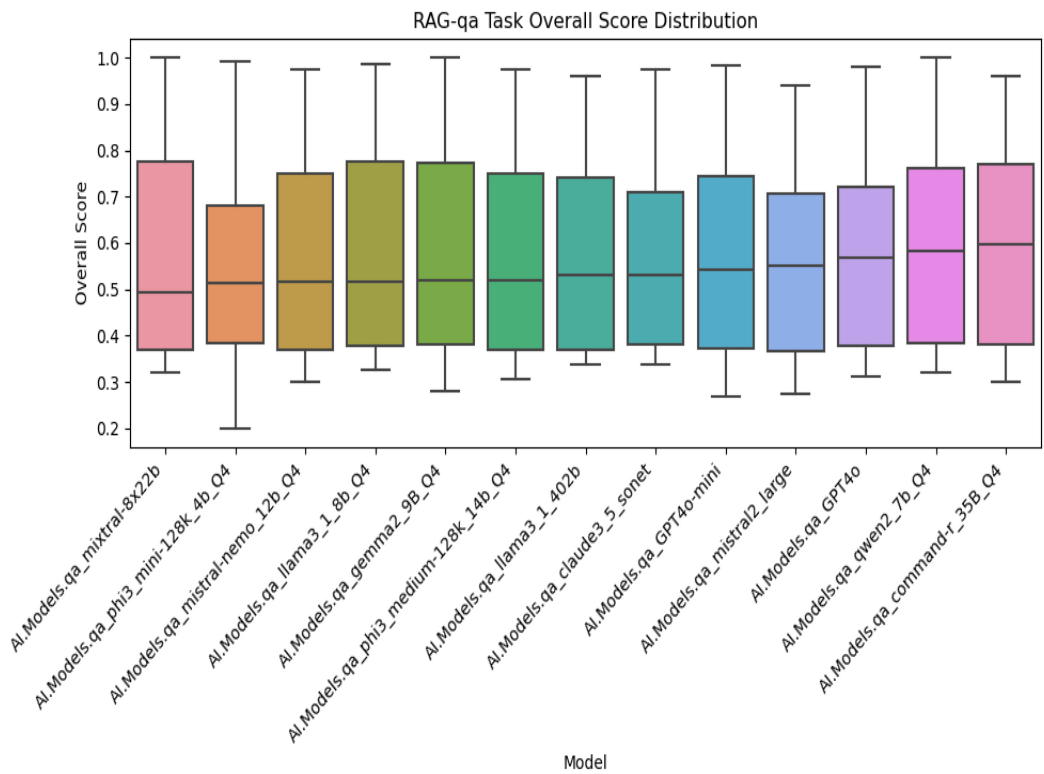


Rag-qa Task Results

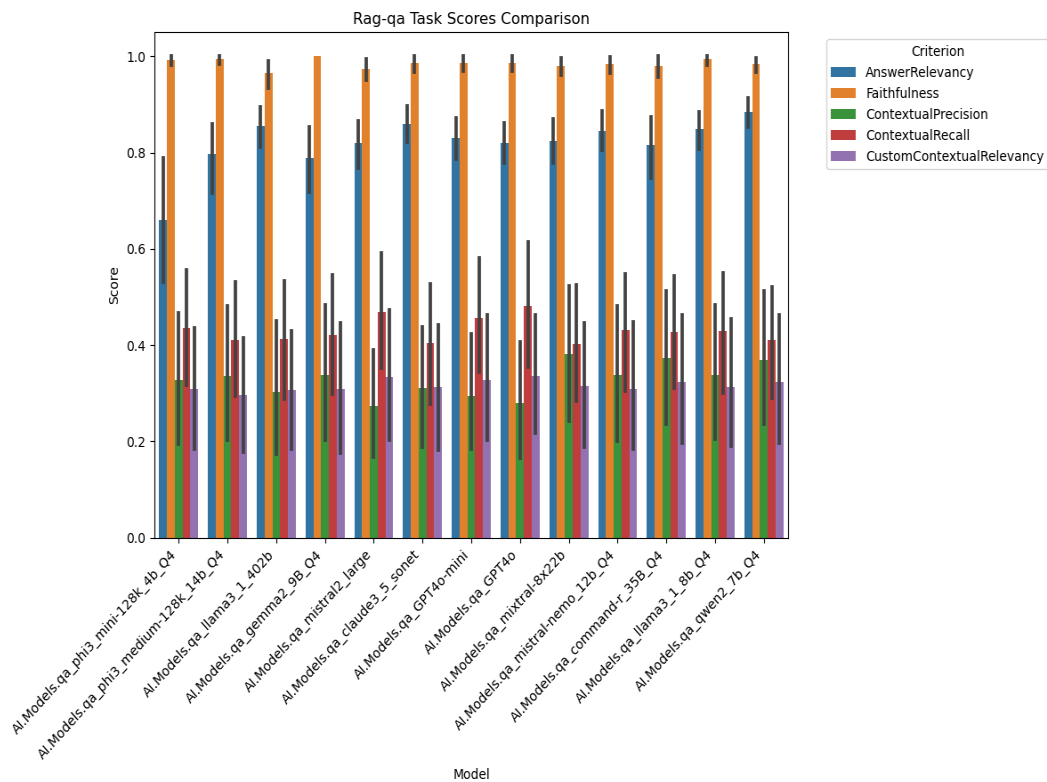




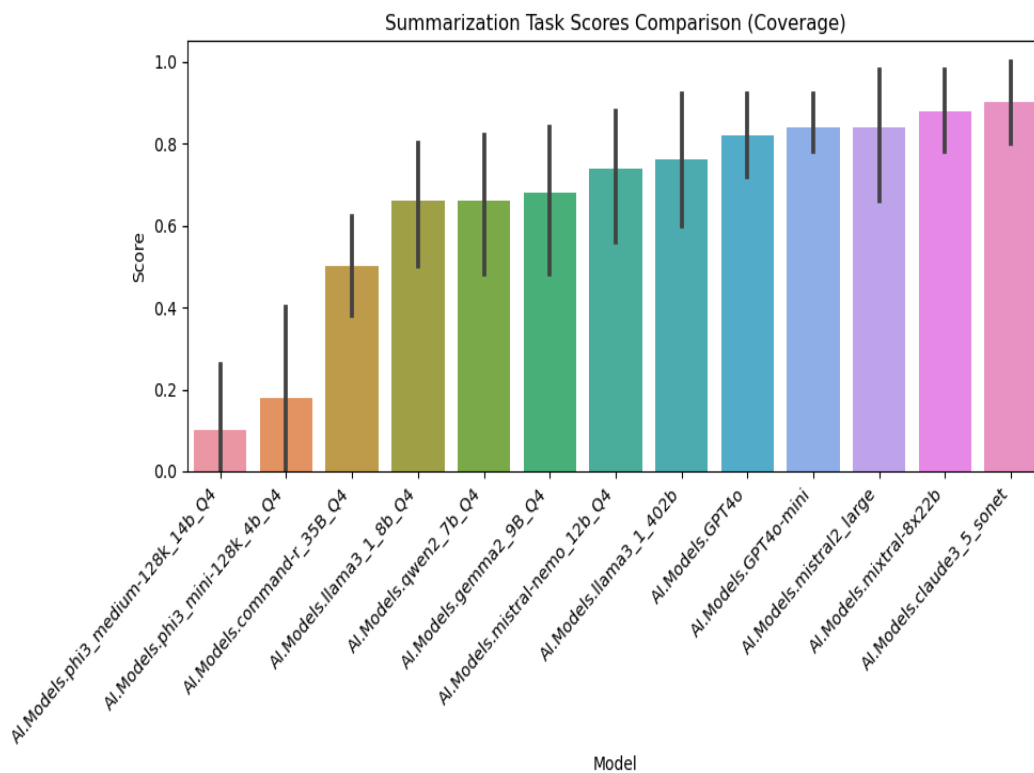
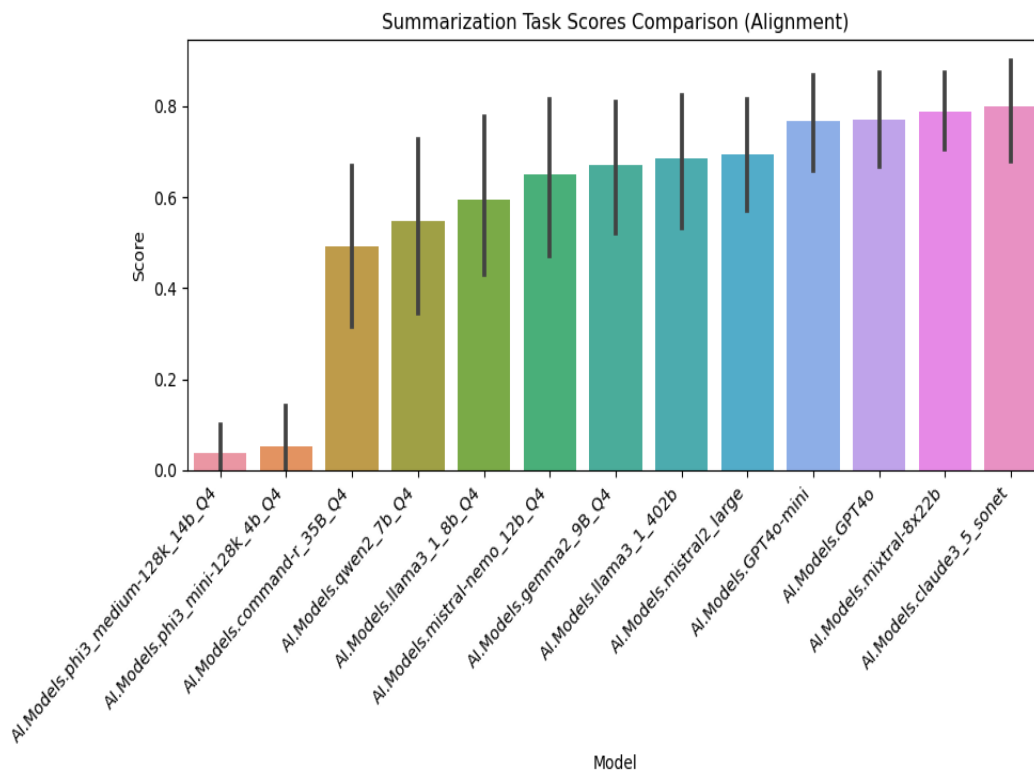
Overall Score Distribution



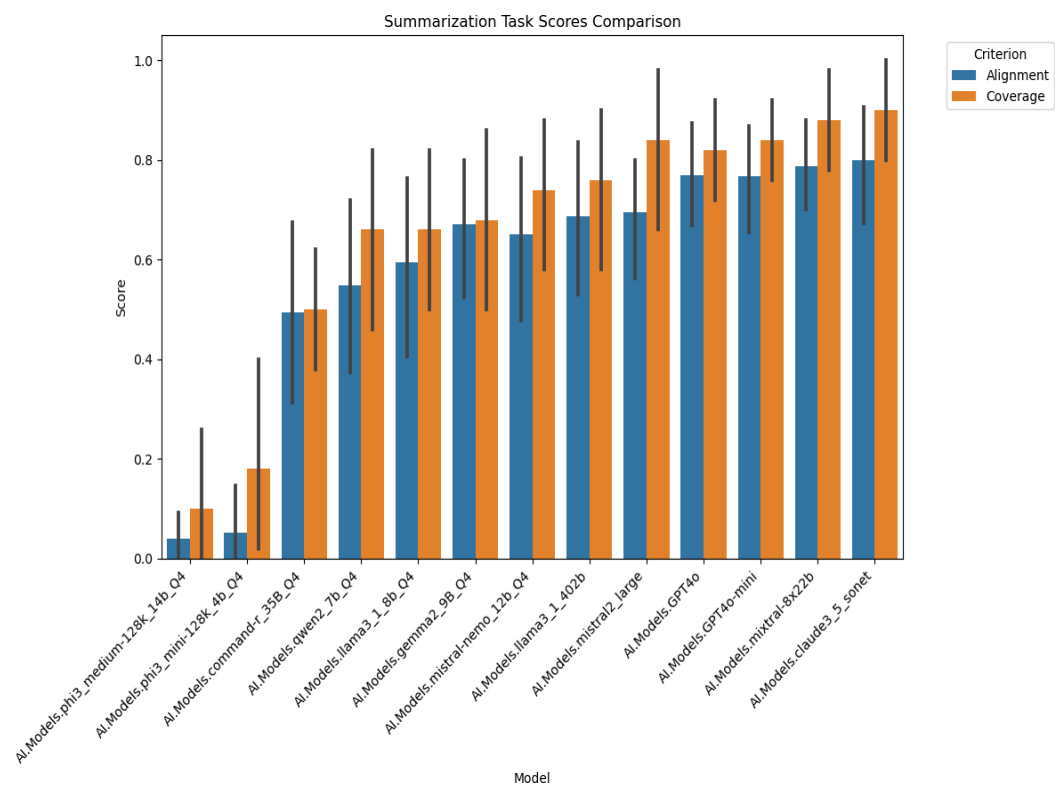
Grouped Results



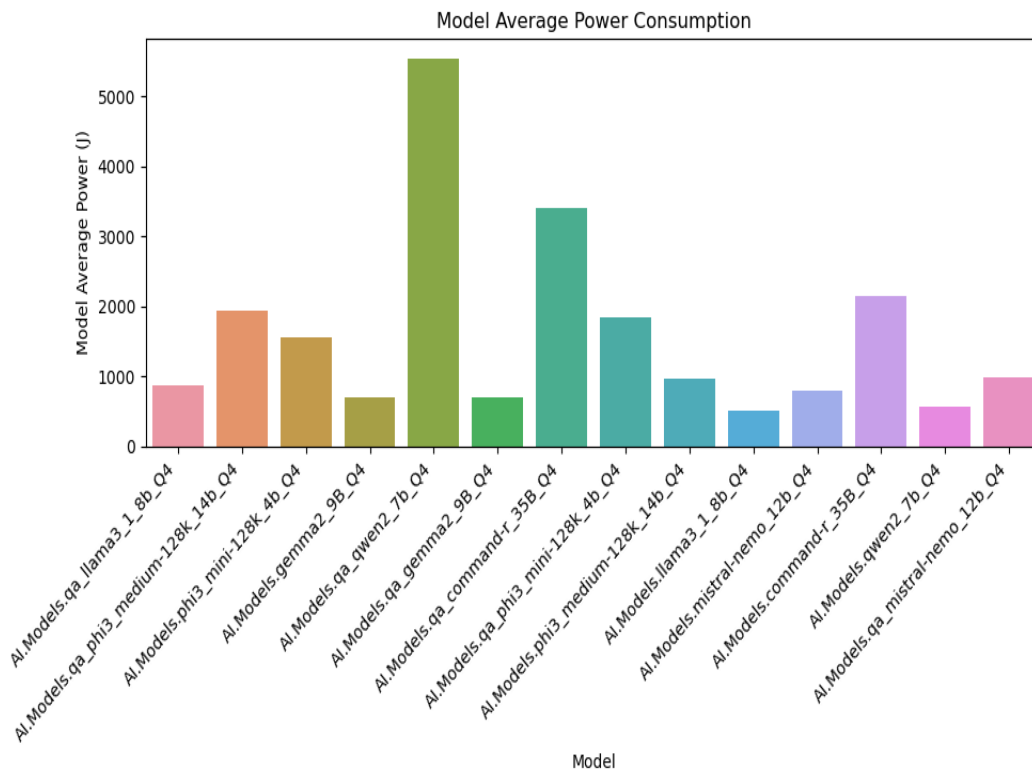
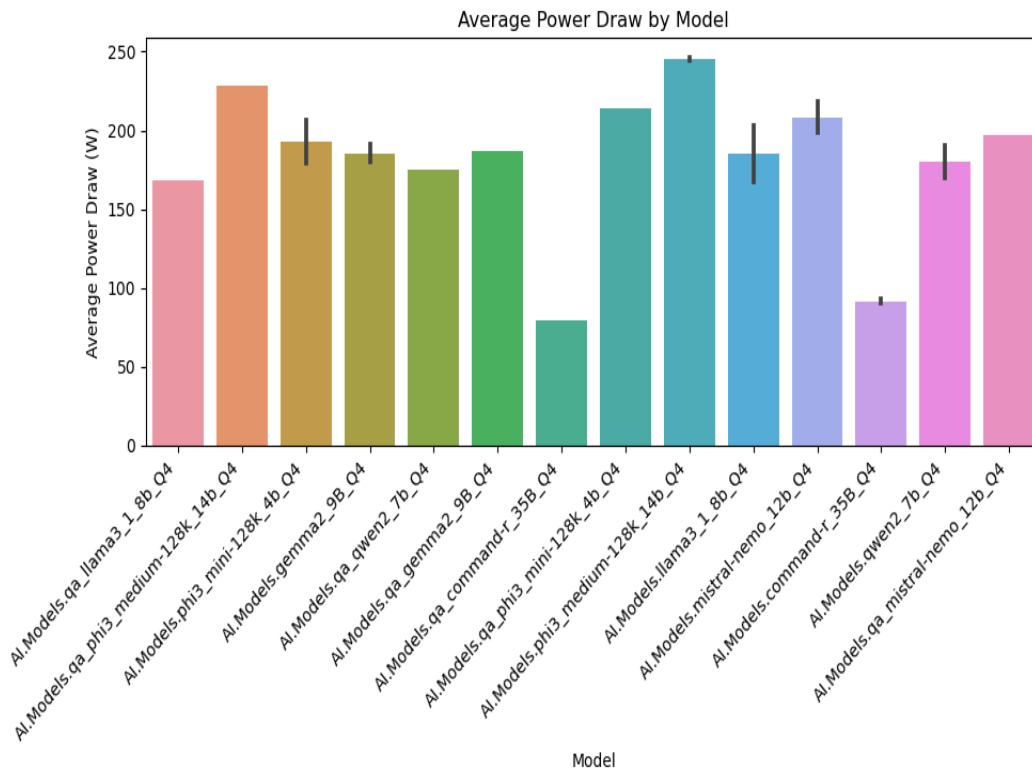
Summarization Task Results

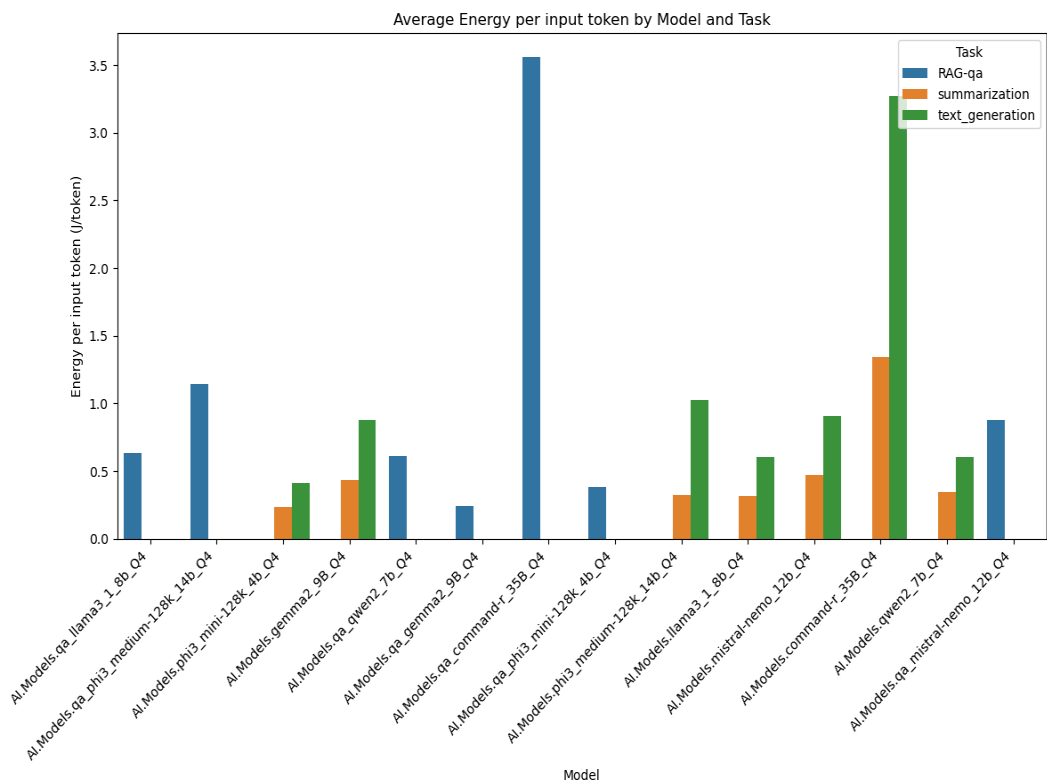
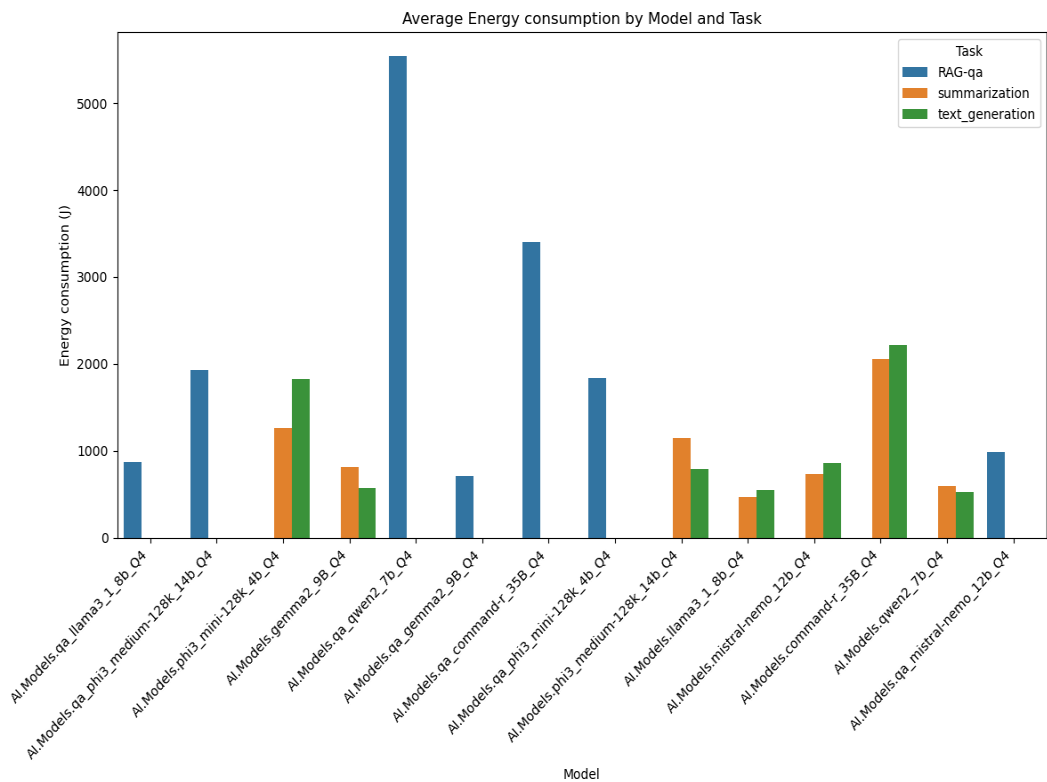


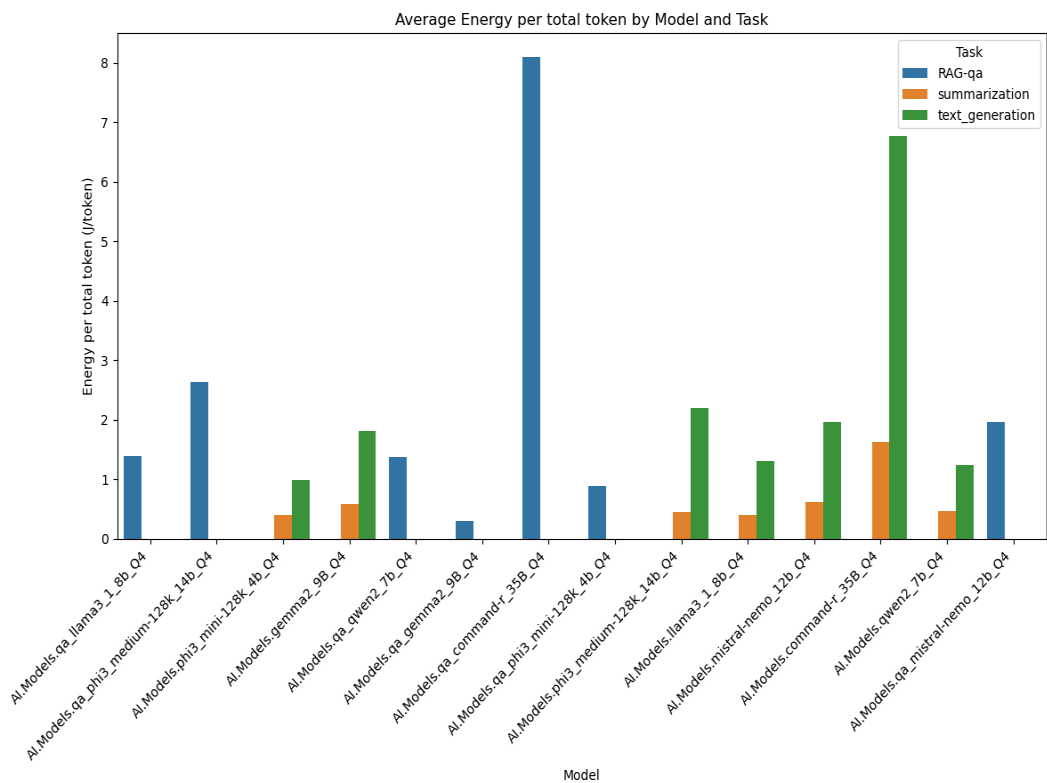
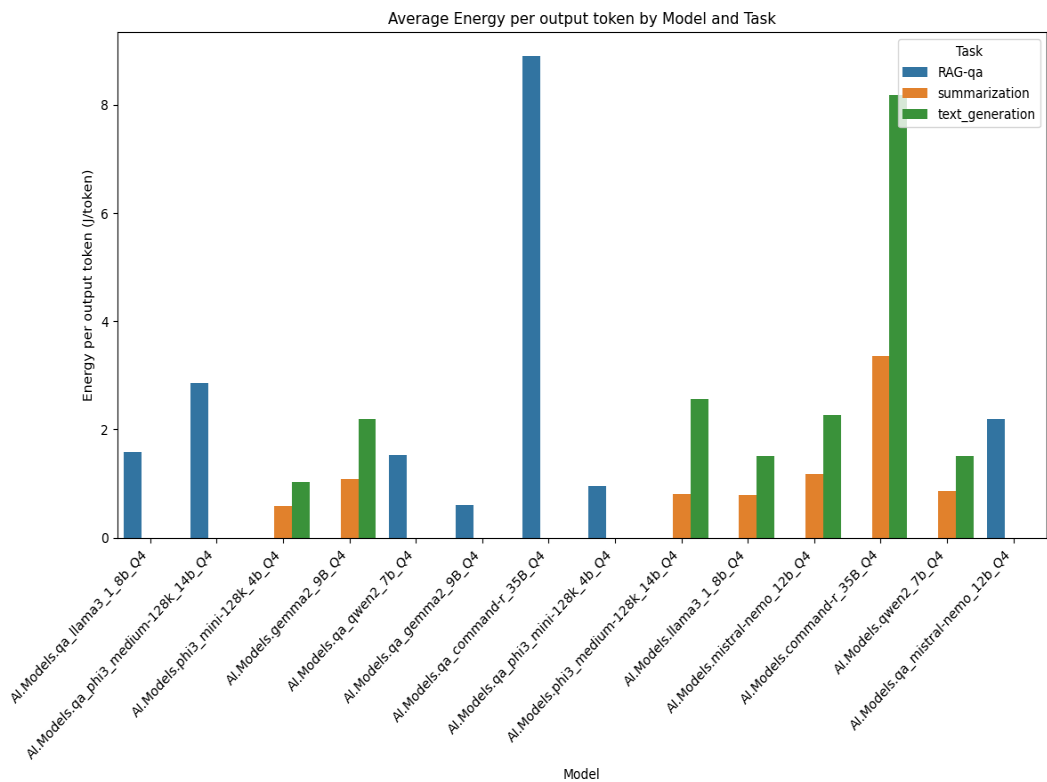
Grouped Results



Power Consumption Results







Evaluation Results

Rag-qa Results

Model: AI.Models.qa_GPT4o

File: qa_001_result.json

Overall_score: 0.4704761904761905

Individual_scores:

AnswerRelevancy: 0.8095238095238095

Faithfulness: 0.9

ContextualPrecision: 0.14285714285714285

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.81 because the response is mostly relevant to the issue of the missing notifications bell in XWiki, but it includes repetitive source links that do not directly address the problem, slightly reducing the overall relevancy.

Faithfulness: The score is 0.90 because the actual output incorrectly claims that the notifications feature requires JavaScript, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.14 because the relevant node, which mentions editing the 'xwiki.properties' file to enable notifications, is ranked seventh. Irrelevant nodes, such as the first node discussing 'installation instructions and prerequisites for extensions', should be ranked lower since they do not address enabling notifications in XWiki.

ContextualRecall: The score is 0.50 because while the concept of enabling notifications is supported by node 7 in the retrieval context, the specific detail about the 'notifications.enabled' setting in 'xwiki.properties' is not found in any of the nodes.

CustomContextualRelevancy: The retrieval context does not mention the 'notifications.enabled' setting in 'xwiki.properties', which is crucial for enabling notifications as stated in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6523504273504274

Individual_scores:

AnswerRelevancy: 0.6923076923076923

Faithfulness: 1.0

ContextualPrecision: 0.5694444444444444

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.69 because the response provides some relevant information about the issue of not receiving notifications in XWiki, but it includes multiple irrelevant source links that do not directly address the question, which lowers 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 0.57 because relevant nodes in the retrieval context, such as the first node explaining 'how to follow a user to receive notifications' and the sixth node mentioning 'different ways to be notified in XWiki', are ranked higher than irrelevant nodes. However, irrelevant nodes like

the second node discussing 'technical filters and settings that do not directly address the issue' and the third node about 'customizing email templates and creating global notification filters' are ranked too closely to relevant nodes, which lowers the score.

ContextualRecall: The score is 0.50 because while some sentences, like those about the 'alert' menu (sentence 4), following a user (sentence 5), and the 'network' tab (sentence 6), are supported by nodes in the retrieval context, others such as enabling notifications in `xwiki.properties` (sentence 2), subscribing to pages or users (sentence 3), and watching a page or wiki (sentence 4) are not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions following a user and watching pages, but lacks details on enabling notifications via `xwiki.properties` or using the alert menu.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9818181818181818

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

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 in XWiki. However, it includes a link to RSS documentation, which is not directly relevant to the specific question asked.

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 states that 'By default, one don't receive notification about one's own activity' and mentions the ability to disable this filter in the 'Advanced filtering options' section of your own notification settings, which directly addresses the expected output. The second node mentions that '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 the expected output. All irrelevant nodes, such as the third node discussing following other users and default watch settings, 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 accurately 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 change this setting in the 'Advanced filtering options' of the user's notification settings.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6142424242424243

Individual_scores:

AnswerRelevancy: 0.9545454545454546

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.75

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.95 because the answer is highly relevant and provides useful information on disabling notifications in XWiki. However, the inclusion of a source link without directly addressing the specific question 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 0.17 because the relevant node in the retrieval context, ranked sixth, 'mentions that users only get notifications for pages they created or updated and discusses notification settings', which is pertinent to disabling notifications. However, several irrelevant nodes are ranked higher, such as the first node that 'discusses following a user and default watch settings, which is not directly related to disabling notifications for pages', and the second node that 'talks about selecting the level of details in email notifications and macros for settings, which does not address disabling notifications'. These irrelevant nodes should be ranked lower 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 discusses automatic page watching, changing settings, and managing custom filters. However, it lacks specific mention of being notified about changes on one's work.

CustomContextualRelevancy: The retrieval context mentions default watch settings and user profile settings but lacks specific details about automatic page watching and deleting notification filters as described in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8666666666666668

Individual_scores:

AnswerRelevancy: 0.7333333333333333

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.73 because the response partially addresses the configuration issue by providing some relevant information. However, it includes multiple irrelevant statements, such as unnecessary links and references to unrelated topics, which prevent 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 1.00 because the relevant node, which includes the property '# oidc.enableUser=true', is ranked first and directly addresses enabling user accounts on first login. All irrelevant nodes, discussing topics like 'domain-based authentication and session cookies', 'resetting passwords', and 'container authentication and LDAP', are correctly ranked lower.

ContextualRecall: The score is 0.60 because while the retrieval context covers the key property '# oidc.enableUser=true' and its purpose as mentioned in sentences 1 and 2 of the expected output, it lacks details on saving the file and restarting XWiki (sentence 4) and the automatic enabling of accounts (sentence 5).

CustomContextualRelevancy: The retrieval context includes the exact information from 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.9326086956521739

Individual_scores:

AnswerRelevancy: 0.9130434782608695

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.91 because the answer largely addresses the question about customizing user profile fields in XWiki. However, it includes a source link and mentions available variables, which do not directly contribute to solving the specific query about mapping additional fields, preventing 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 1.00 because the relevant node in the retrieval context, which mentions 'oidc.user.mapping=myxproperty1=\${oidc.user.subject}' directly relating to mapping additional user profile fields from the OpenID Connect provider to XWiki user properties, is ranked first. All irrelevant nodes, such as those discussing access to various claims, group synchronization, authentication management, and general OpenID Connect guides, are ranked lower, demonstrating perfect contextual precision.

ContextualRecall: The score is 0.75 because the 3rd node in the retrieval context supports the mapping of user profile fields and provides examples, but it does not mention the need to save the xwiki.properties file and restart XWiki to apply the changes, which is why the score is not perfect.

CustomContextualRelevancy: The retrieval context contains detailed instructions on customizing user profile fields using the 'oidc.user.mapping' property in the 'xwiki.properties' file, matching the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.7766666666666667

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.8333333333333334

ContextualRecall: 0.25

CustomContextualRelevancy: 0.8

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 on providing a focused and accurate 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 0.83 because the first two nodes in the retrieval context are relevant and directly address the process of configuring group synchronization by discussing 'oidc.groups.mapping'. However, the third node, which is ranked higher than it should be, primarily provides links to installation instructions and dependencies, which are not directly related to configuring group synchronization. Additionally, the fourth node focuses on authentication management and OpenID Connect endpoints, which are not directly relevant to the task at hand. These irrelevant nodes being ranked higher than the last relevant node discussing 'oidc.groups.claim' slightly lowered the score.

ContextualRecall: The score is 0.25 because only a few specific details, such as 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', mentioned in nodes 1 and 7 in the retrieval context, align with the expected output. However, many general instructions and example codes in the expected output are not supported by the retrieval context.

CustomContextualRelevancy: The retrieval context includes information about configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', which are key elements in the expected output. However, it lacks explicit instructions about locating and editing the 'xwiki.properties' file and restarting the XWiki instance.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6304761904761904

Individual_scores:

AnswerRelevancy: 0.9523809523809523

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.7

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant and mostly addresses the question about restricting file uploads to images or PDFs of a specific size. However, it includes a link to a general guide on XWiki basic concepts, which may not directly address the specific query, slightly affecting its 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 0.20 because the relevant node, which discusses 'setting limits on the maximum size of attachments' and mentions the 'XWikiPreferences document', is ranked fifth. This relevant information should be ranked higher than the first node, which is about 'linking to external files and images', unrelated to configuring upload restrictions or size limits.

ContextualRecall: The score is 0.70 because while sentences related to configuring mimetypes and maximum attachment size (supported by nodes 5 and 6 in the retrieval context) are well-supported, other sentences lack direct support from the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restriction in xwiki.properties, but lacks specifics on 10MB limit and allowed mimetypes configuration.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.7103030303030302

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 0.8333333333333334

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.82 because the answer mostly addresses how to deny script rights to a space administrator, but it includes multiple irrelevant source links that do not directly contribute to answering the question.

Faithfulness: The score is 0.83 because the actual output incorrectly claims that denying script rights to space administrators overrides their inherited administrative permissions, while the retrieval context does not support this assertion.

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 ninth, discusses various topics such as 'programming rights and their implications', 'lists of rights and their properties', 'tabular view of rights', 'comment, edit, and delete rights', 'script right and its default status', 'code documentation for access checks', 'converting rights and excluding classes and pages', 'checkAccess() method and access violations', and 'potential issues with configurable applications and administration access', but none address the specific issue of denying script rights to a space administrator.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports the expected output, with clear evidence from the nodes in the retrieval context confirming that admin rights cannot be denied. Great job!

CustomContextualRelevancy: The retrieval context confirms that rights implied by admin cannot be denied, aligning closely with the expected output. However, it lacks a direct statement about XWiki not supporting this, slightly reducing completeness.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7427350427350428

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 1.0

ContextualPrecision: 0.5694444444444444

ContextualRecall: 0.875

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.77 because the output partially addresses the question about configuring permissions in XWiki, but it includes multiple irrelevant statements with source links that do not directly answer the configuration 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 0.57 because the first node in the retrieval context is relevant, discussing 'rights inheritance and administration features' pertinent to configuring access control in XWiki. However, the second node is irrelevant, focusing on 'the parent/child relationship and nested spaces', which does not directly relate to setting view and edit permissions for teams. Additionally, the third node is also irrelevant, discussing 'string formatting and OpenID Connect properties', which is unrelated to the input question. These irrelevant nodes are ranked higher than other relevant nodes,

such as the sixth node, which explains 'XWiki's permission system, including setting view and edit rights at different levels', and the eighth node, which provides 'basic rules about setting rights at different levels in XWiki'.

ContextualRecall: The score is 0.88 because most of the expected output is supported by the nodes in retrieval context, particularly node 6 which aligns with the majority of the steps and outcomes related to access control. However, there is a lack of information regarding the adjustment of rights for additional teams or spaces, which slightly impacts the score.

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and inheriting access rights, which aligns with setting wiki-wide and space-level rights in the expected output. However, it lacks specific details about creating groups for each team and setting rights specifically for 'XWikiAllGroup' and team-specific edit rights.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.37

Individual_scores:

AnswerRelevancy: 0.85

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses the issue of the notification bell not being available, but includes repeated irrelevant statements about providing a source link instead of directly resolving the problem.

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 'custom notification filters and their behavior' but does not address enabling notifications in the wiki settings. The second node lists 'various ways to receive notifications in XWiki' but does not mention enabling notifications through 'xwiki.properties'. The third node provides 'links to notification preferences' but does not address enabling notifications in the wiki settings. The fourth node provides a 'description of the Notifications Application' but does not discuss enabling notifications in the wiki settings. The fifth node outlines 'various notification settings and options' but does not mention enabling notifications in the wiki settings. The sixth node discusses 'administrator settings for notifications' but does not specifically mention enabling notifications through 'xwiki.properties'. The seventh node focuses on 'email notification settings' and does not address 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 node(s) in the retrieval context, as there is no mention of enabling notifications or setting 'notifications.enabled' in 'xwiki.properties'.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.37894736842105264

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

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.

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 'links and images related to various references and do not address notification settings or issues', while the fifth node 'provides information about 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 linked to any nodes in the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.36470588235294116

Individual_scores:

AnswerRelevancy: 0.8235294117647058

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the question about enabling notifications for one's own actions. However, the presence of multiple irrelevant source links 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 'themes, internationalization, accessibility, and scalability', which are unrelated to notification settings. The second node is about 'administering wiki features like users and rights', which does not pertain to notification settings. The third node covers 'importing office documents, icon themes, and syntax highlighting', none of which relate to notification settings. The fourth node is about 'customizing themes and panels', which is not relevant to enabling notifications for one's own actions. The fifth node discusses 'color themes and access rights', which do not address 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, indicating a complete mismatch between the expected content and the provided context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.3333333333333333

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the question about disabling notifications, but includes several irrelevant links that do not directly help with the issue, 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'. Similarly, the second node is about 'moving attachments', and the third node lists 'display and content features', none of which pertain to the input query about disabling notifications. This pattern continues throughout all nodes, 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, indicating a complete lack of relevant information.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.3955555555555555

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 0.8333333333333334

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant and provides useful information about configuring XWiki for automatic user activation with OpenID Connect. However, it includes an incomplete statement that slightly detracts from the overall clarity and completeness.

Faithfulness: The score is 0.83 because the actual output incorrectly claims that the OpenID Connect authentication service configuration should be in the 'xwiki.cfg' file with a specific entry, 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 instance authentication and session cookie

configuration', which is unrelated to 'automatically activating user accounts on first login with OpenID Connect'. The second node covers 'container authentication and Active Directory integration', not the configuration needed for 'automatic account activation with OpenID Connect'. Subsequent nodes similarly fail to address the specific configuration for automatic user activation, such as the 'oidc.enableUser' property.

ContextualRecall: The score is 0.20 because only the general topic of configuring OpenID Connect in XWiki is supported by the 5th node in the retrieval context, while specific details like the 'oidc.enableUser' property, its configuration, and the process of saving and restarting are not covered.

CustomContextualRelevancy: The retrieval context does not contain information about modifying the xwiki.properties file or the oidc.enableUser property for automatic user account activation.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.5133333333333333

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.87 because the response is mostly relevant and addresses the customization of user profile fields in XWiki, but it includes source links that do not directly answer 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 'various authentication methods including OpenID Connect' but lacks information on 'customizing user profile fields or mapping them in the xwiki.properties file.' The second node focuses on 'domain-based authentication and session cookie configuration,' which is unrelated to the input query. Subsequent nodes continue to miss the topic of 'customization of user profile fields' or 'oidc.user.mapping,' leading to a complete mismatch with the input.

ContextualRecall: The score is 0.40 because while there are nodes in the retrieval context that relate to customizing user profile fields with OpenID Connect (node 3) and using 'oidc.user.mapping' (node 7), the example mappings, variable replacements, and instructions to save and restart XWiki are not explicitly supported by the retrieval context.

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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.44000000000000006

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

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 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 focuses on general OpenID Connect provider configuration and endpoints, which does not address group synchronization setup. The second node discusses container authentication and other methods, lacking details on configuring group synchronization between XWiki and an OpenID Connect provider. The third node is about domain-based instance authentication and session cookies, unrelated to group synchronization setup. The fourth node provides a general description of OpenID Connect authentication and mentions automatic synchronization, but lacks specific instructions for group synchronization setup. The fifth node lists dependencies for the OpenID Connect extension, which is not relevant to setting up group synchronization. The sixth node mentions group synchronization but does not provide detailed instructions or configuration steps necessary for the expected output. The seventh node provides a brief mention of group synchronization using a claim, but lacks detailed configuration steps or examples needed for the expected output.

ContextualRecall: The score is 0.20 because only a few aspects of the expected output, such as setting properties in ``xwiki.properties`` and defining group mappings using ``oidc.groups.mapping``, are supported by the nodes in the retrieval context. Most instructions, like locating and editing the ``xwiki.properties`` file, configuring specific properties, and restarting the XWiki instance, are not covered in the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.5708452950558213

Individual_scores:

AnswerRelevancy: 0.8421052631578947

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5454545454545454

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.84 because the response mostly addresses the question about configuring upload restrictions for images or PDFs, but it includes irrelevant links to pages about exports, document lifecycle, and XWiki syntax, which 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 0.17 because the relevant node in the retrieval context, which provides information on setting the maximum size of an attachment and mentions MIME type restrictions, is ranked sixth. This is lower than several 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 configuration topics like WebDAV and PDF export',

which should be ranked lower.

ContextualRecall: The score is 0.55 because while the steps related to configuring the maximum upload size (sentences 2-5) are well-supported by the 6th node in the retrieval context, the initial instructions about MIME type configuration (sentence 1) are not supported by any node in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting maximum 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 described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.3666666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 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 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 does not address denying script rights. The second node is about 'template providers and page creation', unrelated to script rights management. The third node contains 'technical code snippets for configuration', not related to rights management. The fourth node involves 'user registration and validation', which is not relevant to denying script rights. The fifth node discusses 'access rights and execution context' but does not specifically address the denial of script rights to a space administrator.

ContextualRecall: The score is 0.00 because the expected output mentions admin rights in XWiki, but none of the nodes in the retrieval context provide information about admin rights or their implications.

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 answer.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.45595238095238094

Individual_scores:

AnswerRelevancy: 0.5714285714285714

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.375

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.57 because the output contains several irrelevant links that do not directly address configuring access rights in XWiki, which is the main focus of the input. However, the score is not lower because there is some relevant information related to access rights configuration.

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

ContextualPrecision: The score is 0.33 because the relevant nodes in the retrieval context, like the third node which 'mentions administrating wiki features like users, groups, and rights, which is directly relevant to configuring access controls for teams,' and the sixth node which 'provides a link to managing users and groups, which is relevant for setting up team-specific access controls in XWiki,' are ranked lower than irrelevant nodes. For example, the first node 'discusses XWiki's database support and efficiency, which is not relevant to configuring access controls for teams,' and the second node 'is about displaying notifications and does not relate to access control configurations in XWiki,' both of which are ranked higher than the relevant nodes.

ContextualRecall: The score is 0.38 because only some aspects of the expected output are supported by the nodes in the retrieval context. Specifically, sentence 1 is supported by node 5 regarding managing groups, sentence 2 is supported by node 3 about administrating rights, and sentence 4 is supported by node 5 regarding nested pages. However, sentences 3, 5, 6, 7, and the introductory and reminder sentences do not have direct matches in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks information on creating groups, setting wiki-wide and area-specific rights, and inheriting rights for subpages, which are key elements in the expected answer.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.8136363636363637

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the issue of the notification bell not being available, 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 0.25 because the relevant node in the retrieval context, ranked 4th, mentions the configuration variable 'notifications.enabled' in 'xwiki.properties', which is directly relevant to enabling notifications. However, it is ranked lower than three irrelevant nodes that discuss 'custom notification filters and their behavior', 'links and descriptions of various notification preferences', and 'filters for notifications and technical pages', none of which address enabling notifications in the wiki settings.

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 explicitly mentions enabling notifications by setting 'notifications.enabled' to true in xwiki.properties, which matches the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.7177777777777777

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 0.7222222222222222

ContextualRecall: 0.6666666666666666

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.90 because the response is highly relevant to the issue of not receiving notifications, but it includes 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 0.72 because the relevant nodes in the retrieval context are generally 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 relate to the activation of notifications. However, the second node, which focuses on 'customizing notification email templates and creating global notification filters', is not directly related to enabling notifications and should be ranked lower. Similarly, the fourth node, discussing 'RSS feeds and grouping notifications', and the fifth node, about 'displaying preferences of the current user about notifications', do not directly address enabling notifications or following users/pages and should be ranked lower than the relevant nodes.

ContextualRecall: The score is 0.67 because while the retrieval context nodes support the general idea of enabling notifications and following entities for updates (as seen in nodes 1 and 3), specific details like setting 'notifications.enabled' in 'xwiki.properties' and viewing followed users in the 'network' tab are missing.

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.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.9436363636363637

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.82 because the answer mostly addresses the question about enabling notifications for personal actions, but includes links to unrelated sources, which slightly detract from its 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 1.00 because the relevant node in the first rank directly addresses the input question 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,' which is exactly what the input is about. All irrelevant nodes, starting from the second node, discuss topics like 'RSS feeds and general notification settings,' 'a list of links and topics related to notifications,' 'various notification preferences macros,' 'notification email settings and developer options,' and 'following users and default watch settings,' none of which specifically address the issue of receiving notifications for one's own actions. Great job on getting it spot on!

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 by default, users do not receive notifications for their own actions and mentions the option to change this setting, aligning closely with the expected output. However, it lacks specific 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: 0.5671428571428571

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.79 because the response addresses the question about disabling notifications, but includes irrelevant links that do not directly help with the task, slightly 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 'various technical aspects of notification filters and settings' but does not address how to disable notifications or 'auto-watched pages'. The second node lists 'different notification preferences and macros' without providing information on disabling notifications. The third node provides links about notifications but lacks specific instructions on disabling them. The fourth node talks about 'getting RSS feeds and general notification settings' without mentioning disabling notifications. Finally, the fifth node gives an overview of notification methods but does not include details on disabling notifications or managing auto-watched pages.

ContextualRecall: The score is 0.75 because while nodes in the retrieval context support the functionality of auto-watching pages and managing notification settings (as seen in nodes 3 and 4), there is no direct mention of being notified of changes on your work, which affects the completeness of the expected output attribution.

CustomContextualRelevancy: The retrieval context mentions auto-watch preferences and notification settings, but lacks details on modifying auto-watch settings and removing filters from 'Filtres personnalisés'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.490000000000000005

Individual_scores:

AnswerRelevancy: 0.95

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant and addresses the main question about configuring OpenID Connect for automatic user activation. However, it includes a mention of LDAP Authenticator, 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 'domain-based authentication and session cookie configuration', which does not address enabling user accounts automatically on first login. The second node provides information about 'OpenID Connect endpoints and headers', but lacks details on user account activation settings. Subsequent nodes continue to focus on unrelated topics such as 'configuration guides', 'dependencies for the OpenID Connect extension', and 'LDAP and Active Directory integration', none of which pertain to the specific issue of user account activation during first login.

ContextualRecall: The score is 0.50 because while the nodes in retrieval context support the mention of 'xwiki.properties' and 'oidc.enableUser' (sentence 1 and 2), they do not directly support the specific instructions for adding or updating properties (sentence 3) or ensuring new user accounts are activated (sentence 5).

CustomContextualRelevancy: The retrieval context lacks any mention of adjusting the xwiki.properties file or the specific property oidc.enableUser, which is crucial for the expected answer.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.5538461538461539

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.8

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.77 because the answer partially addresses the question about customizing user profile fields in XWiki, but it includes repetitive irrelevant statements that provide source links without directly 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 'various authentication methods and configurations but does not mention customizing user profile fields or mapping them in xwiki.properties.' The second node is about 'domain-based authentication and session cookies, which is unrelated to mapping user profile fields from OpenID Connect.' The third node lists 'endpoints, bypassing OpenID Connect, and group synchronization, but does not address user profile field mapping.' All nodes fail to address the specific topic of mapping user profile fields from OpenID Connect in XWiki.

ContextualRecall: The score is 0.80 because most of the expected output aligns with the information in the 2nd node in the retrieval context, such as references to 'xwiki.properties', 'oidc.user.mapping', and the process of saving and restarting XWiki. However, the retrieval context lacks a specific example of mapping properties, which affects the completeness of the match.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6576923076923077

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.5384615384615384

CustomContextualRelevancy: 0.5

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, 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 0.25 because the relevant node in the retrieval context, which 'directly addresses group synchronization, mentioning the use of the 'oidc.groups.mapping' property', is ranked fourth. However, it should be ranked higher than the irrelevant nodes. For example, the first node 'discusses authentication management and OpenID Connect endpoints, but does not address group synchronization or the specific configuration steps needed for it', making it less relevant to the input question.

ContextualRecall: The score is 0.54 because while the 4th node in the retrieval context provides detailed information on configuring group synchronization, including properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', it lacks specific instructions found in the expected output, such as locating and editing the 'xwiki.properties' file, replacing placeholder values, and saving changes.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.34545454545454546

Individual_scores:

AnswerRelevancy: 0.7272727272727273

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.73 because the answer partially addresses the question about configuring file upload restrictions, but it includes multiple irrelevant source links that do not directly relate to the specific task of restricting uploads to images or PDFs under 10 MB.

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 talks about 'various XWiki applications like tour guides and administration', which also does not pertain to file upload settings. Similarly, the third node is about 'customizing PDF export styles using CSS and XSL', which is not relevant to the topic of file upload restrictions. All subsequent nodes continue this trend of irrelevance, discussing topics like 'trusted domains and PDF export architecture', 'ODT export and PDF export methods', 'XWiki macros and their functionalities', and 'references and images', none of which address the input question regarding file upload configurations.

ContextualRecall: The score is 0.00 because none of the expected output details, such as file size restrictions, MIME type configurations, or specific steps for setting upload limits, are supported by any nodes in the retrieval context.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.720952380952381

Individual_scores:

AnswerRelevancy: 0.9047619047619048

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.90 because the response is mostly relevant to the question about managing script rights for a space administrator. However, it includes source links about internationalization and the XWiki RESTful API, which are not directly related to 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 0.50 because the second and fourth nodes in the retrieval context are relevant as they discuss 'script rights can be denied' and 'explains script rights, including the ability to deny them', which are pertinent to the question about denying script rights to an administrator. However, the first node, which is about 'language preferences and translations in XWiki', and the third node, which is about 'interacting with the XWiki RESTful API', are both irrelevant to the question and should be ranked lower than the relevant nodes.

ContextualRecall: The score is 1.00 because the expected output is perfectly supported by the 5th node in the retrieval context, showing a complete match without any discrepancies. Great job!

CustomContextualRelevancy: The retrieval context mentions that certain rights like 'programming' and 'admin' are not deniable, but it does not explicitly state that admin rights cannot be refused, which is the main point of the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.3444444444444444

Individual_scores:

AnswerRelevancy: 0.7222222222222222

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.72 because the response addresses the main question about configuring access rights in XWiki, but it includes multiple irrelevant source links that do not contribute to answering 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 question about configuring access rights for team spaces in XWiki. For example, the first node provides 'metadata about a specific page in XWiki', which is not relevant, and the second node 'explains how to modify resources using XML and HTTP methods', which is unrelated to setting access permissions. Similarly, the third node 'discusses comment management', which does not pertain to configuring team space access. 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, as all key concepts and steps mentioned are not covered in the retrieval context nodes.

CustomContextualRelevancy: The retrieval context does not contain any information related to creating groups, defining rights, or inheriting rights, which are the key components of the expected answer.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.8175000000000001

Individual_scores:

AnswerRelevancy: 0.6875

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.69 because the response partially addresses the question about authenticating users with access tokens, but it includes multiple 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. Keep up the great work!

ContextualPrecision: The score is 0.50 because relevant nodes, such as the second node, which 'provides information on configuring authorized applications', and the fourth node, which 'describes the requirements for JWT tokens', are ranked lower than irrelevant nodes. For example, the first node

'discusses LDAP authentication settings', which is unrelated to token-based authentication using JWTs, and should be 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, showcasing a perfect alignment. Great job!

CustomContextualRelevancy: The retrieval context includes details about JWT token requirements, authorized applications configuration, and the use of Ed25519 keys, matching most of the expected output. However, it lacks explicit mention of the `xwiki.authentication.authclass` property setting.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.3111111111111111

Individual_scores:

AnswerRelevancy: 0.5555555555555556

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.56 because the output includes multiple irrelevant source links that do not directly address the question about authenticating users with access tokens, which lowers the relevancy. However, the score is above average, indicating that some relevant information is still 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 'model validation strategies in XWiki', which is unrelated to 'token-based authentication'. The second node talks about 'attachment validation and mime type restrictions', which do not pertain to 'authentication with access tokens'. The third node provides information about 'the admin UI for model validation', which is not relevant to 'user authentication using access tokens'. The fourth node contains a 'code snippet for setting up a registration field', which is unrelated to 'token-based user authentication'. Finally, the fifth node describes 'steps for installing UI extensions and performing upgrades', which are not relevant to 'the authentication process using access tokens'.

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 token-based authentication or JWT, which are key elements in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.3142857142857143

Individual_scores:

AnswerRelevancy: 0.5714285714285714

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.57 because the response includes some relevant information about authenticating users with access tokens, but it is diluted by irrelevant links to documentation and extensions that do not directly address 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 question. The first node discusses 'upgrading XWiki' without mentioning user authentication with access tokens. The second node covers 'user rights and registration' but lacks information on authentication using access tokens. The third node, although about security, states that the 'Authentication API is not yet implemented', which is not helpful for authenticating users with access tokens. The fourth node focuses on 'managing rights' and does not provide details on access token authentication. Finally, the fifth node lists 'different rights within XWiki' without addressing 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.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_GPT4o-mini

File: qa_001_result.json

Overall_score: 0.5333333333333334

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0.30952380952380953

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.86 because the main content effectively addresses the issue with the 'Bell' icon in 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. Great job on maintaining perfect alignment with the retrieval context!

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 for extensions,' which does not relate to enabling notifications in XWiki. Similarly, the second node 'addresses issues with configurable applications and administration access,' not related to the bell icon issue. However, the third node provides 'information about the Notifications Application and mentions receiving notifications in the top bar,' which is relevant. Additionally, the seventh node 'mentions that administrators can configure notification settings and edit the xwiki.properties file,' which is also pertinent to the input query.

ContextualRecall: The score is 0.50 because while the retrieval context does mention configuring notifications in a general sense (node 7), it lacks specific details about setting 'notifications.enabled' in 'xwiki.properties', which is crucial for full alignment with the expected output.

CustomContextualRelevancy: The retrieval context does not mention the 'notifications.enabled' setting in 'xwiki.properties', which is crucial for enabling notifications as outlined in the expected output.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_002_result.json

Overall_score: 0.8028571428571428
Individual_scores:
AnswerRelevancy: 0.7
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 0.7142857142857143
CustomContextualRelevancy: 0.6
Reasons:

AnswerRelevancy: The score is 0.70 because the output contains relevant information about notification issues in XWiki, but it is cluttered with irrelevant headings 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 1.00 because the relevant node in the retrieval context, which 'explains how to follow a user and receive notifications about events triggered by them', is ranked first, ensuring that the most pertinent information is prioritized. Great job on getting it right!

ContextualRecall: The score is 0.71 because several sentences in the expected output, such as those about watching pages or wikis, using the alert menu, following users, and viewing the network tab, are well-supported by nodes in the retrieval context. However, the instructions about enabling notifications in `xwiki.properties` and subscribing to pages or users are not directly supported by any nodes in the retrieval context, leading to a lower score.

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

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_003_result.json

Overall_score: 0.9846153846153847
Individual_scores:
AnswerRelevancy: 0.9230769230769231
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 1.0
CustomContextualRelevancy: 1.0
Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant and provides useful information on enabling notifications for one's own actions in XWiki. However, it includes an unrelated statement about RSS features, which 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 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant nodes. Great job on 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 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6286507936507936

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 0.8571428571428571

ContextualPrecision: 0.29166666666666663

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and provides useful information on how to disable notifications in XWiki. However, it includes a source link that does not directly address the question, slightly impacting the score.

Faithfulness: The score is 0.86 because the actual output incorrectly claims the existence of an 'Advanced filtering options' section to disable notifications about your own activity or hide minor events, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.29 because the relevant nodes in the retrieval context are not ranked as high as they should be. The fourth node is relevant as it 'mentions that administrators can change users' notification settings and discusses auto-watched mode,' but it is preceded by three irrelevant nodes. The first node is irrelevant because it 'discusses following a user and default watch settings, but does not mention how to disable notifications or manage notification settings in the user profile.' Similarly, the second node 'talks about the details of changes in notifications and macros for settings, but does not provide information on disabling notifications or managing notification settings.' The third node 'focuses on enabling and disabling the email feature and customizing live email notifications,' which is not directly related to the input query. Additionally, the sixth node, which is relevant as it 'discusses the settings of the notifications menu and mentions that users can enable or disable notification types,' is ranked lower than several irrelevant nodes.

ContextualRecall: The score is 0.75 because the 5th node in the retrieval context supports most of the expected output by discussing automatic page watching, user profile settings, and custom filters. However, it does not specifically mention being notified about changes on your work, which affects the score.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8777777777777779

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.89 because the answer is mostly relevant and addresses the question about configuring OpenID Connect in XWiki. However, it includes an irrelevant statement about LDAP authentication, which slightly detracts from the focus on 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 first node in the retrieval context is directly relevant, mentioning the configuration line '# oidc.enableUser=true', which is essential for enabling user accounts on first login. All irrelevant nodes, discussing topics like 'domain-based authentication and session cookies' or 'resetting passwords and various authentication mechanisms', are ranked lower, ensuring perfect precision. Great job on the ranking!

ContextualRecall: The score is 0.60 because while the 1st node in the retrieval context provides information about adjusting the 'xwiki.properties' file and the specific property to enable user accounts on first login, it lacks details about saving the file, restarting the XWiki instance, and the automatic enabling of new user accounts.

CustomContextualRelevancy: The retrieval context contains the commented line '# oidc.enableUser=true', which matches the expected answer's instruction to uncomment and set this property to true. However, the retrieval context does not explicitly mention saving the file and restarting the XWiki instance.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.925

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.875

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 1.0

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. Great job!

Faithfulness: The score is 0.88 because the actual output suggests a need to configure user profile rights using the 'oidc.user.ownProfileRights' property, whereas the retrieval context indicates that the default setting is 'edit', implying no configuration is necessary.

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 examples of mapping, is ranked first. All other nodes discuss topics like access tokens, group synchronization, authentication management, and general OpenID Connect information, which are not directly related to customizing user profile fields.

ContextualRecall: The score is 0.75 because the 3rd node in the retrieval context supports the customization of user profile fields and provides examples of mapping user properties, but it does not mention the need to save the 'xwiki.properties' file and restart XWiki to apply changes.

CustomContextualRelevancy: The retrieval context includes the exact information about customizing user profile fields using the 'oidc.user.mapping' property in the 'xwiki.properties' file, matching the expected output precisely.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.7866666666666667

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.8333333333333334

ContextualRecall: 0.4

CustomContextualRelevancy: 0.7

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 0.83 because the first two nodes in the retrieval context are relevant, providing information on the 'oidc.groups.mapping' property and group synchronization process. However, the third node, ranked higher than it should be, discusses 'installation instructions and dependencies', which are not directly related to configuring group synchronization. Similarly, the fourth node, which discusses 'authentication management and endpoints', should be ranked lower as it is not relevant to the group synchronization process. Despite these misplacements, the relevant nodes are mostly ranked higher, contributing to the current score.

ContextualRecall: The score is 0.40 because while nodes in the retrieval context provide specific details about properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', they do not support the general instructions and introductory statements found in the expected output.

CustomContextualRelevancy: The retrieval context includes information about configuring the 'oidc.groups.claim' and 'oidc.groups.mapping' properties, but lacks explicit instructions for locating and editing the 'xwiki.properties' file, as well as saving and restarting XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6555555555555556

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.7777777777777778

CustomContextualRelevancy: 0.3

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. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.20 because the relevant node, which mentions 'setting limits on the maximum size of an attachment and configuring it in XWikiPreferences', is ranked fifth. The first four nodes discuss unrelated topics such as 'linking to external files and images', 'general information about

XWiki's structure', 'adding attachments without addressing restrictions', and 'trusted domains and PDF export customization', which should be ranked lower than the relevant node.

ContextualRecall: The score is 0.78 because while nodes in the retrieval context provide detailed steps for configuring maximum upload size (Step 2, sentences 3-7), the introductory and explanatory parts of Step 1 (sentences 1-2) lack direct support from the retrieval context nodes.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific details about configuring xwiki.properties and setting size to 10MB.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.7436363636363637

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.82 because the main content effectively addresses the question about denying script rights, 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. 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 '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 properties', but does not mention the inability to deny script rights to a space administrator. Similarly, the third node provides 'a tabular view of rights' but does not discuss the specific issue of denying script rights to a space administrator. All subsequent nodes also fail to address the specific question about denying script rights to a space administrator, leading to a score of 0.00.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the nodes in the retrieval context, confirming that rights implied by admin cannot be denied. Great job!

CustomContextualRelevancy: The retrieval context states that admin rights imply other rights and are not deniable, aligning with the expected output that rights implied by admin cannot be denied. However, the specific mention of 'unfortunately not supported in XWiki' is not directly found in the context.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.6286666666666666

Individual_scores:

AnswerRelevancy: 0.76

Faithfulness: 1.0

ContextualPrecision: 0.20833333333333331

ContextualRecall: 0.875

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.76 because while the response addresses the configuration of permissions in XWiki, it includes several irrelevant source links that do not directly contribute to solving the problem, thus 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 0.21 because the relevant nodes in the retrieval context, such as the sixth node discussing 'XWiki's permission system, including setting view and edit rights at different levels,' and the eighth node outlining 'the basic rules for setting rights in XWiki, including wiki-wide and page-level rights,' are ranked lower than several irrelevant nodes. For instance, the first node discusses 'content organization and rights inheritance in XWiki but does not provide specific steps for configuring view and edit permissions for teams,' and the second node explains 'the parent/child relationship and nested spaces concept but does not address how to set view and edit permissions for teams,' both of which should be ranked lower than the relevant nodes.

ContextualRecall: The score is 0.88 because most of the expected output sentences are supported by nodes in the retrieval context, particularly the 6th node which covers various aspects of access control and rights management. However, there is a lack of specific information regarding adjusting rights for additional teams or spaces, which affects the score.

CustomContextualRelevancy: The retrieval context mentions rights inheritance and administration features, but lacks details on creating groups, setting specific rights, and ensuring teams can only edit their own spaces.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.3666666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the issue of the notification bell not being available, but includes irrelevant information about sources and documentation links that do not directly relate to solving 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 to the input. Each node, from the first to the seventh, discusses topics such as 'custom notification filters', 'different ways to receive notifications', 'notification preferences macros', 'Notifications Application', 'various settings and developer options', 'configuring notifications for administrators', and 'email notification settings', none of which address enabling notifications in the wiki settings. Therefore, they should all be ranked lower.

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 setting `notifications.enabled` to `true`, which is the key information in the expected answer.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.37894736842105264

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.89 because the response mostly addresses the issue of not receiving notifications, but it includes irrelevant source links that do not pertain 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 to various projects,' which do not relate to notifications. The second node also includes 'image links and references,' lacking information on notifications. The third node continues with 'image links and references,' offering no relevant details. The fourth node is a 'continuation of image links and references,' without mentioning notifications. The fifth node discusses 'permissions and a security module,' unrelated to 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 contain any information related to enabling notifications or following pages and users as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.3684210526315789

Individual_scores:

AnswerRelevancy: 0.8421052631578947

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.84 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 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 query about enabling notifications for user activity. The first node discusses 'customization of page templates, menus, and CSS files', which is unrelated. The second node provides information on 'administrating wiki features', not pertaining to notifications. The third node describes 'features like Office Importer and Font Awesome Icon Theme', which are not relevant. The fourth node discusses 'themes, panels, and presentation settings', unrelated to the query. Lastly, the fifth node is about 'color

themes and access rights', which does not relate to notifications or user activity 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, as there is no mention of notifications, user activity, or disabling filters in the user's notification settings.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.3125

Individual_scores:

AnswerRelevancy: 0.5625

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.56 because the output contains several irrelevant URLs that do not directly address the question about disabling notifications, which lowers the score. However, the presence of some relevant information about notifications keeps the score above the midpoint.

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 example, the first node discusses 'attaching files' and the second node talks about 'moving attachments', neither of which relate to notification settings or disabling notifications. Additionally, the third node is about 'displaying content and macros', which is also unrelated 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 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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.5

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about configuring XWiki for automatic user activation with OpenID Connect, without any irrelevant information. Great job!

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 query. The first node discusses 'domain-based authentication and session cookies', which is unrelated to enabling user accounts automatically upon first login with OpenID Connect. The second node mentions OpenID Connect but lacks information about automatic user activation. Subsequent nodes also fail to address the specific configuration needed for automatic user activation, focusing instead on topics like 'installation', 'general information', 'properties and endpoints', 'dependencies', and a 'table of contents'.

ContextualRecall: The score is 0.50 because while the retrieval context mentions the `xwiki.properties` file and the `oidc.enableUser` setting in node 5, it lacks specific instructions and the effect of the configuration change, leading to partial alignment with the expected output.

CustomContextualRelevancy: The retrieval context lacks any mention of the xwiki.properties file or the oidc.enableUser property, which are essential for automatic user account activation as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.5422222222222223

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.4

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.94 because the response is mostly relevant and addresses the question about mapping additional user profile fields from an OpenID Connect provider to XWiki user properties. However, the statement 'Zusätzliche Felder in der Antwort.' is vague and does not directly contribute to solving the issue, preventing a perfect 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 0.17 because the relevant node in the retrieval context, ranked sixth, mentions the `xwiki.properties` file and includes details about OpenID Connect endpoints, which is relevant to configuring user profile fields. However, the first five nodes are irrelevant as they discuss topics such as container authentication, domain-based authentication, and general information about the OpenID Connect Authenticator without mentioning user profile field customization or the `xwiki.properties` file. These irrelevant nodes are ranked higher than the relevant node, leading to a lower score.

ContextualRecall: The score is 0.40 because while the 7th node in the retrieval context supports the mention of 'xwiki.properties' and 'oidc.user.mapping', key details such as the example mapping syntax, placeholder replacement instructions, and the need to save and restart XWiki are missing.

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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.4852380952380953

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.22619047619047616

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the output 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 0.23 because the relevant nodes in the retrieval context, which discuss 'group synchronization' and the 'oidc.groups.mapping' property, are ranked lower (sixth and seventh) than the irrelevant nodes. The irrelevant nodes, ranked from first to fifth, focus on topics like 'authentication management and OpenID Connect endpoints', 'container authentication', 'domain-based authentication', 'OpenID Connect authentication', and 'dependencies for the OpenID Connect extension', which do not address group synchronization with OpenID Connect.

ContextualRecall: The score is 0.20 because only a few elements from the expected output, such as the 'oidc.groups.claim' and 'oidc.groups.mapping' properties, are supported by nodes in the retrieval context. Most of the detailed steps and configurations mentioned in the expected output, like setting up group synchronization, finding and editing the 'xwiki.properties' file, and specific configuration instructions, 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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.609792663476874

Individual_scores:

AnswerRelevancy: 0.7368421052631579

Faithfulness: 1.0

ContextualPrecision: 0.26666666666666666

ContextualRecall: 0.5454545454545454

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.74 because the response partially addresses the input question about upload restrictions, but the presence of multiple 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 0.27 because the relevant nodes in the retrieval context, specifically the fifth node which includes sections on 'Size Limit' and 'Mimetype Restriction', and the sixth node which explains setting limits on attachment size in the XWikiPreferences document, are ranked lower than several irrelevant nodes. The first four nodes discuss topics such as 'trusted domains and PDF export customization', 'ODT and PDF export features', 'WebDAV and PDF export', and 'PDF and RTF export issues', which are unrelated to configuring attachment restrictions, yet they are ranked higher

than the relevant nodes.

ContextualRecall: The score is 0.55 because while several specific instructions related to editing configurations and navigating the interface are well-supported by nodes in the retrieval context, key details about MIME type configuration and file size limitations are not directly mentioned in any nodes.

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 10 MB limit.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.37142857142857144

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.86 because the answer is mostly relevant and addresses the question about denying script rights to a space administrator. However, it includes irrelevant source links that do not directly contribute to answering the question, preventing a higher score.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment 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 'validation patterns and script execution', which is not related to denying script rights to an admin in XWiki. The second node is about 'page creation templates' and does not address rights management or script permissions. The third node describes a 'custom right check class and configuration', but it does not mention script rights or their denial. The fourth node involves 'field validation and registration processes', which are unrelated to script rights or administrative permissions. Finally, the fifth node discusses 'access rights and execution contexts', but it does not specifically address the inability to deny script rights to an admin in XWiki.

ContextualRecall: The score is 0.00 because the expected output sentence about admin rights in XWiki does not correspond to any nodes in the retrieval context, as none of them mention admin rights or their implications.

CustomContextualRelevancy: The retrieval context does not contain any information about XWiki not supporting the denial of rights implied by admin rights, which is the main point of the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.33076923076923076

Individual_scores:

AnswerRelevancy: 0.6538461538461539

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.65 because the response partially addresses the question about configuring permissions in XWiki, but includes several irrelevant links that do not directly relate to the task of setting permissions, which 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 'XWiki's database support and programming features,' which do not relate to configuring access controls for teams. The second node is about the 'Notifications Application,' which is unrelated to setting access permissions for teams in XWiki. The third node covers the 'Administration Application,' which involves user and group management but does not specifically address configuring team access permissions. Subsequent nodes continue to focus on unrelated topics such as 'Model Validation UI,' 'Nested Pages and subwikis,' and 'LDAP group configurations,' none of which provide the needed information for configuring team-specific access controls in XWiki.

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

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.8192307692307693

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.85 because the response is mostly relevant to the issue of the notification bell not being available, but it includes links that are not directly related to solving 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, which mentions the 'notifications.enabled' property in 'xwiki.properties', is ranked fourth. The first three nodes are irrelevant as they discuss 'custom notification filters and their behavior', 'various notification preferences', and 'notification filters and settings', none of which address enabling notifications in the wiki.

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 explicitly mentions enabling notifications by setting 'notifications.enabled' to true in xwiki.properties, which matches the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.7187301587301588

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0.7222222222222222

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.86 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 are 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 0.72 because relevant nodes are generally ranked higher, such as the first node discussing 'Enable and disable the extension' and the third node about 'Follow a user', which are pertinent to receiving notifications. However, the second node, which focuses on 'email frequency and template customization', is ranked higher than it should be, as it does not directly address enabling notifications. Similarly, the fourth node on 'RSS feeds and macro settings' and the fifth node about 'displaying notification preferences' are also ranked higher than necessary, affecting the score.

ContextualRecall: The score is 0.71 because while several nodes in the retrieval context support the expected output, such as enabling notifications and following users or pages (nodes 1 and 3), there are missing details about specific configurations like 'xwiki.properties' and the 'réseau' tab, which are not mentioned in the retrieval context.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8618181818181819

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.91 because the response is highly relevant to the question about enabling notifications for one's own actions, 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, 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 about not receiving notifications for one's own activity. This ensures that irrelevant nodes, which discuss unrelated notification settings, lists of links, and other notification preferences, are ranked lower.

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 about disabling the filter in 'Options de

filtrage avancées' lacks direct support from the retrieval context nodes.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.5354545454545455

Individual_scores:

AnswerRelevancy: 0.7272727272727273

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.73 because the response partially addresses the question about disabling notifications, but includes irrelevant links that do not directly help with 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 'various notification filters and settings' but lacks information on 'automatic page watching' or disabling such notifications. The second node lists 'different notification preferences and macros' without addressing how to stop notifications for 'automatically watched pages'. The third node provides 'a list of notification settings and options' but does not cover 'automatic following of pages'. The fourth node talks about 'RSS feeds and notification settings' without mentioning 'automatic watching of pages'. Lastly, the fifth node outlines 'various notification methods in XWiki' but fails to address 'automatically watched pages'.

ContextualRecall: The score is 0.75 because while nodes in the retrieval context support the idea of default notification settings and accessing notification settings (as seen in nodes 2 and 4), they do not explicitly mention being notified of changes on one's own work, which affects the completeness of the expected output.

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'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.5046153846153847

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant to configuring OpenID Connect in XWiki, but it includes an irrelevant link about LDAP authentication, 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 'domain-based authentication and session cookies', which is unrelated to enabling user accounts automatically upon first login using OpenID Connect. The second node provides information about 'OpenID Connect endpoints and configurations', but does not mention enabling user accounts automatically upon first login. The third node describes 'the OpenID Connect authenticator and its installation', but does not address the specific configuration needed to enable user accounts automatically. The fourth node lists 'dependencies for the OpenID Connect extension', which is not relevant to the configuration needed for automatic user account activation. The fifth node is about 'LDAP and Active Directory integration', which is not related to OpenID Connect or the automatic activation of user accounts. The sixth node provides 'general information about the OpenID Connect authenticator', but does not discuss the specific configuration for automatic user account activation. The seventh node discusses 'container authentication and LDAP', which are unrelated to the OpenID Connect configuration needed for automatic user account activation.

ContextualRecall: The score is 0.40 because while there is a connection between the expected output and the nodes in retrieval context regarding the `xwiki.properties` file and the property `oidc.enableUser`, several key instructions and assurances, such as updating the property, saving the file, restarting the instance, and automatic activation of user accounts, are not supported by the nodes in retrieval context.

CustomContextualRelevancy: The retrieval context mentions the xwiki.properties file but does not include specific details about setting oidc.enableUser to true for automatic user activation.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.48000000000000001

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.80 because the answer largely addresses the question about customizing user profile fields in XWiki using OpenID Connect, but it is not higher due to repeated irrelevant statements providing source links 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 and various authentication methods', which does not mention customizing user profile fields. The second node is about 'domain-based instance configuration and session cookies', unrelated to mapping user profile fields. The third node lists features of the OpenID Connect Authenticator without providing information on mapping user profile fields. Subsequent nodes similarly fail to address the specific topic of mapping user profile fields in XWiki.

ContextualRecall: The score is 0.40 because while there is some relevant information in the nodes in retrieval context about OpenID Connect and xwiki.properties, specific details about mapping custom properties, replacing placeholders, and the steps to save and restart XWiki are missing.
CustomContextualRelevancy: The retrieval context mentions OpenID Connect and xwiki.properties but lacks specific details about customizing user profile fields or using oidc.user.mapping.
Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_027_result.json

Overall_score: 0.57
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 0.25
ContextualRecall: 0.3
CustomContextualRelevancy: 0.3
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. Great job on maintaining perfect alignment with the retrieval context!
ContextualPrecision: The score is 0.25 because the relevant node, which 'directly addresses group synchronization, mentioning the use of claims and mappings,' is ranked fourth. However, it is preceded by three irrelevant nodes that discuss 'authentication management and endpoints,' 'container authentication,' and 'domain-based instance authentication and session cookies,' none of which relate to group synchronization. This lowers the score as the relevant information should be ranked higher than these irrelevant contexts.
ContextualRecall: The score is 0.30 because while the 4th node in the retrieval context supports configuring group synchronization using properties like `oidc.groups.claim` and `oidc.groups.mapping`, many key instructions such as locating the `xwiki.properties` file, opening it with a text editor, and specific examples for configuration are missing in the retrieval context.
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, and saving changes in xwiki.properties.
Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_028_result.json

Overall_score: 0.352
Individual_scores:
AnswerRelevancy: 0.76
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:
AnswerRelevancy: The score is 0.76 because the answer partially addresses the question about configuring file upload restrictions in XWiki, but it includes multiple 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 'language support and translations in XWiki', which is unrelated to file upload restrictions. Similarly, the second node is about 'various XWiki applications and extensions', not addressing file upload settings. Subsequent nodes continue to focus on unrelated topics such as 'customizing PDF export with CSS and XSL', 'trusted domains for redirections', and 'ODT export and PDF export features', none of which pertain to file upload settings or restrictions.

ContextualRecall: The score is 0.00 because none of the expected output content, such as file size restrictions, MIME type configurations, or specific configuration steps, is supported by the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about configuring MIME types or setting maximum attachment sizes, which are key details in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.6368421052631579

Individual_scores:

AnswerRelevancy: 0.6842105263157895

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.68 because the output contains some relevant information about script rights, but it includes multiple 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 0.50 because relevant nodes in the retrieval context, such as the second node which 'explains the implications of programming rights and mentions that script rights are part of the administrative rights, which cannot be denied once granted', and the fourth node which 'provides details on script rights and how they are managed, indicating that script rights are allowed by default unless explicitly revoked', are ranked lower than irrelevant nodes. The first node, discussing 'language preferences and translations in XWiki', is unrelated to script rights and should be ranked lower. Similarly, the third node about 'interacting with the XWiki RESTful API' is not relevant to script rights or administrative permissions and should also be ranked lower.

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 does not mention that administrator rights cannot be refused, which is the key information in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.35238095238095235

Individual_scores:

AnswerRelevancy: 0.7619047619047619

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.76 because the response addresses the main question about configuring access rights in XWiki, but includes several irrelevant source links that do not directly relate to the 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 0.00 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. The second node discusses 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, which is not relevant to configuring access rights for spaces in XWiki. These irrelevant nodes are ranked higher than any relevant nodes, 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, as the context lacks any mention of the specific steps, rights settings, or group configurations described in the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information related to creating groups, setting rights, or inheriting rights in XWiki as outlined in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.823076923076923

Individual_scores:

AnswerRelevancy: 0.6153846153846154

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.62 because the response partially addresses the question about authenticating users with access tokens, but it includes multiple irrelevant source links that do not directly contribute to answering the query.

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 0.50 because relevant nodes about 'Authorized Applications' and JWT token generation, which are pertinent to token-based authentication, are ranked lower than irrelevant nodes. For instance, the first node focuses on 'LDAP authentication settings', which does not relate to token-based methods, while the second node is relevant as it discusses 'Authorized Applications'. Similarly, the fourth node, which is relevant due to its focus on 'JWT token generation and claims', is ranked lower than the first node. This misordering affects the precision score.

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 achieving a perfect match!

CustomContextualRelevancy: The retrieval context fully supports the expected answer by detailing the use of JWT tokens, required claims, Ed25519 key signing, and configuration steps for the Token-based

authentication extension.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.35

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.75 because the main content addresses the question about authenticating users with access tokens, 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. 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 'model validation strategies in XWiki', the second node is about 'attachment validation and configuration in XWiki', the third node provides 'administrative information about model validation in XWiki', the fourth node describes 'a script for setting up a registration field in XWiki', and the fifth node is about 'installing a UI in XWiki'. None of these contexts relate to authenticating users with access tokens, leading to the lowest possible 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.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.2690909090909091

Individual_scores:

AnswerRelevancy: 0.5454545454545454

Faithfulness: 0.8

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.55 because the response includes some relevant information about authenticating users with access tokens, but it is diluted by multiple irrelevant links to documentation and extension pages that do not directly address the question.

Faithfulness: The score is 0.80 because the actual output inaccurately claims that the security module of XWiki handles user authentication, while the retrieval context clarifies that the Authentication API is not implemented yet and the old core code is still in use.

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 talks about 'features related to users and rights in XWiki', but does not mention 'token-based authentication'. The third node covers 'security modules and APIs', but explicitly states that 'the Authentication API is not yet implemented', which does not support 'the use of access tokens for authentication'. The fourth node is about 'managing rights and permissions in XWiki', which is not directly related to 'authenticating users with access tokens'. Finally, the fifth node provides 'a table of rights constants and their descriptions', which does not relate 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 matched to any nodes in the retrieval context, indicating a complete lack of relevant information.

CustomContextualRelevancy: The retrieval context lacks any information about token-based authentication or JWT, which are central to the expected answer.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_claude3_5_sonnet

File: qa_001_result.json

Overall_score: 0.3833333333333333

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant and provides useful information on fixing the missing bell icon in XWiki. However, it includes source links that do not directly address the issue, 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 question. The first node discusses 'installation instructions and prerequisites', which do not help in fixing the bell issue. The second node describes 'the functionality of the notifications menu and the bell icon' but fails to address how to enable notifications or resolve the issue. The third node lists 'dependencies for the notifications extension', which does not relate to the problem at hand. Finally, the fourth node talks about 'overriding notification templates', which is unrelated to fixing the bell icon issue. Thus, irrelevant nodes are ranked higher than any relevant information, 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, as there is no mention of enabling notifications or the 'notifications.enabled' setting in 'xwiki.properties'.

CustomContextualRelevancy: The retrieval context lacks any mention of enabling notifications via the 'notifications.enabled' setting in 'xwiki.properties'.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.7528571428571429

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8

ContextualPrecision: 0.75

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.5

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 0.80 because the actual output incorrectly claims that nothing is watched by default in XWiki 15.5+, while the retrieval context clarifies that for new users, nothing is watched by default, but all application types are enabled.

ContextualPrecision: The score is 0.75 because the first and fourth nodes in the retrieval context are relevant, addressing how to follow a user for notifications and various ways to receive notifications in XWiki. However, the second node discusses filters and settings without directly addressing the issue of not receiving notifications, the third node talks about RSS feeds without ensuring notifications are received, and the fifth node is about the 'What's New' application, which is unrelated to the notification setup or troubleshooting. These irrelevant nodes should be ranked lower than the relevant ones.

ContextualRecall: The score is 0.71 because several sentences in the expected output are well-supported by nodes in the retrieval context, such as watching a page or wiki (supported by node 2), using the alert menu (supported by node 4), and following a user (supported by node 1). However, some key details like enabling notifications via `xwiki.properties` and subscribing to pages or users are not covered in the nodes, leading to a less than perfect score.

CustomContextualRelevancy: The retrieval context covers enabling notifications and following users, but lacks details on setting `notifications.enabled` to `true` in `xwiki.properties` and watching pages or wikis.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.975

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses how to enable notifications for one's own actions in XWiki, but it includes source links that do not directly answer the question, 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 1.00 because all relevant nodes in the retrieval context are ranked higher than 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,' which directly addresses the expected output. The second node mentions, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' which aligns with the expected output. Great job on getting the rankings spot on!

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

CustomContextualRelevancy: The expected answer is fully supported by the retrieval context, which states that by default, one does not receive notifications about one's own activity unless explicitly targeted, and this can be changed 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: 0.6925925925925925

Individual_scores:

AnswerRelevancy: 0.9629629629629629

Faithfulness: 1.0

ContextualPrecision: 0.7

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant and mostly addresses the question about disabling notifications in XWiki. However, the statement 'Understanding the Default Behavior' is slightly vague and doesn't directly contribute to the solution, preventing 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 0.70 because the first node in the retrieval context is relevant, providing information about accessing 'settings' to change notification settings. However, the second node, which is about 'following users and default watch settings,' and the third node, discussing 'filtering notifications and hidden pages,' are not directly related to disabling notifications, yet they are ranked higher than the fifth relevant node that discusses 'administrators changing users' notification settings and auto-watching behavior.'

ContextualRecall: The score is 0.50 because while nodes in the retrieval context support the idea of automatic page watching and changing settings in the user profile (related to sentences 1 and 3), they do not explicitly mention notifications about changes on one's work or deleting existing notification filters (related to sentences 2 and 4).

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8666666666666668

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.83 because the response largely addresses the question about configuring OpenID Connect authentication in XWiki, but includes irrelevant information about LDAP and other authentication methods, 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 '# oidc.enableUser=true', is correctly ranked first. This directly addresses the configuration needed to enable user accounts upon their first login, as required by the input. The irrelevant nodes, discussing topics like domain-based instance authentication, general OpenID Connect information, container authentication, and LDAP, are ranked lower, ensuring they do not interfere with the relevant information.

ContextualRecall: The score is 0.60 because while the retrieval context (node 1) provides detailed information on adjusting the 'xwiki.properties' file and setting the 'oidc.enableUser' property, it lacks details on saving the file, restarting XWiki, and the automatic enabling of new user accounts upon first login.

CustomContextualRelevancy: The retrieval context contains the commented-out property '# oidc.enableUser=true', matching the expected answer's requirement to uncomment and set this property to 'true'. However, it lacks explicit instructions to save the file and restart the XWiki instance.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.96

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.8

CustomContextualRelevancy: 1.0

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, ranked first, directly addresses the input by explaining how to 'associate non standard properties coming from the OpenID Connect provider with the XWiki user' and provides an example of mapping. The irrelevant nodes, which discuss topics like 'scopes and response types', 'endpoints and HTTP methods', 'group synchronization', and 'in-wiki configuration', are correctly ranked lower as they do not relate to customizing user profile fields in XWiki.

ContextualRecall: The score is 0.80 because most of the expected output is supported by the 3rd node in the retrieval context, which covers customizing user profile fields and provides examples of mapping. However, the retrieval context lacks information about saving and restarting XWiki to apply changes, which affects the score.

CustomContextualRelevancy: The retrieval context contains the information about customizing user profile fields using the 'oidc.user.mapping' property in the 'xwiki.properties' file, including examples matching the expected answer.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8569230769230771

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9230769230769231

ContextualPrecision: 1.0

ContextualRecall: 0.46153846153846156

CustomContextualRelevancy: 0.9

Reasons:

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

Faithfulness: The score is 0.92 because the actual output incorrectly states the process of adding 'xwiki_groups' to userinfo claims, whereas the retrieval context clarifies that group synchronization is enabled by adding the claim 'xwiki_groups' to the property 'oidc.userinfoclaims'.

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 group synchronization, such as specifying the 'oidc.groups.claim' property, mapping XWiki groups using 'oidc.groups.mapping', and enabling synchronization by adding claims to 'oidc.userinfoclaims'. The irrelevant nodes, ranked fourth and fifth, focus on general configuration and authentication management, which are not directly related to the process of configuring group synchronization. Great job on getting everything in the right order!

ContextualRecall: The score is 0.46 because while some nodes in the retrieval context support configuring properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', many specific instructions such as locating and editing the xwiki.properties file, replacing placeholders, and restarting XWiki are missing.

CustomContextualRelevancy: The retrieval context contains detailed information on configuring 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6485918003565063

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.7272727272727273

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses the question about configuring attachment restrictions, but includes irrelevant details such as a link to a general guide and advice on clearing the browser's cache, which are not directly related 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, ranked third, provides information about setting the maximum size of an attachment and mentions the

XWikiPreferences document, which is directly related to configuring attachment size limits. However, irrelevant nodes are ranked higher, such as the first node, which discusses linking to external files and backlinks indexation, and the second node, which provides an overview of XWiki's structure and features, neither of which address attachment upload restrictions or size limits.

ContextualRecall: The score is 0.73 because while several sentences regarding the configuration of mimetypes and maximum attachment size are supported by nodes in the retrieval context (particularly nodes 3 and 4), introductory and example sentences lack direct support from the 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.7066666666666667

Individual_scores:

AnswerRelevancy: 0.5333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.53 because the output contains multiple irrelevant statements that do not directly address the question about denying script rights to a space administrator, such as providing source links and historical settings. However, it still partially relates to the topic, which is why the score is above zero.

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 programming rights and script execution but does not address the denial of script rights to a space administrator. The second node provides a list of rights and their default states but does not explain the denial of script rights to a space administrator. The third node explains the 'Script' right and its default settings but does not mention denying script rights to a space administrator. The fourth node contains code snippets related to access checks but does not discuss 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 the expected output is perfectly supported by the node 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 answer that rights implied by admin cannot be denied in XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.6985507246376812

Individual_scores:

AnswerRelevancy: 0.8260869565217391

Faithfulness: 1.0

ContextualPrecision: 0.29166666666666663

ContextualRecall: 0.875

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.83 because the response largely addresses the question about configuring permissions in XWiki, but it is not higher due to the inclusion of multiple irrelevant source links that do not directly assist with the configuration task.

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 several irrelevant nodes. The first node discusses 'the history and structure of content organization in XWiki,' which is not related to configuring team-specific permissions. The second node explains 'the parent/child relationship and the concept of Nested Pages,' which also does not address the input question. The third node is about 'string formatting and OpenID Connect user properties,' unrelated to the task. The relevant node, ranked fourth, provides 'information on XWiki's permission system, mentioning the ability to set view and edit rights,' which is pertinent to the input. Another relevant node, ranked sixth, 'outlines the basic rules of XWiki's permission system,' which is also relevant. The presence of these relevant nodes contributes to the score, but their lower ranking compared to irrelevant nodes prevents a higher score.

ContextualRecall: The score is 0.88 because most of the expected output sentences are well-supported by nodes in the retrieval context, such as nodes 4 and 6 aligning with key steps like creating groups and setting rights. However, there's a lack of information in the retrieval context regarding 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 details about creating groups for each team and setting rights specifically for 'XWikiAllGroup' or individual team spaces as outlined in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.5111111111111111

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant to the issue of the notification bell not being available, 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. Everything aligns perfectly with 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' to enable notifications, and should be ranked higher. However, it is ranked lower than five irrelevant nodes that discuss topics like 'custom notification filters', 'Alert menu', 'different ways to receive notifications', 'various notification settings', and 'customizing notification emails', none of which address enabling notifications in the wiki settings.

ContextualRecall: The score is 0.50 because while the retrieval context nodes mention configuring notifications and editing the xwiki.properties file (supporting sentence 1), they do not specifically mention the `notifications.enabled` setting or setting it to `true` (unsupporting sentence 2).

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3888888888888889

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

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 doesn't 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 0.00 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 enabling notifications in XWiki.' The second node 'consists of image links and references, which do not provide any information about notifications settings or configurations.' The third node 'includes image links and project references, offering no relevant information about notification settings 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 related to enabling notifications or following pages and users as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.37142857142857144

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.86 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 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. 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 '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 includes 'information about various extensions like Office Importer and Font Awesome Icon Theme, none of which relate to notifications or user activity settings.' These irrelevant nodes should be ranked lower than relevant nodes.

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 notification settings or user activity filters mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.35454545454545455

Individual_scores:

AnswerRelevancy: 0.7727272727272727

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.77 because the response mostly addresses how to disable notifications, but it includes multiple irrelevant URLs that do not directly help with the issue, 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. 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 and managing attachments in XWiki', which is unrelated to managing notifications. The second node is about 'moving attachments', not about notification settings. The third node lists 'various macros and their functions in XWiki', none of which pertain to notification settings. The fourth node covers 'administrating wiki features' without mentioning notification settings. The fifth node discusses 'permissions and security modules', which are not relevant to managing notifications. Finally, the sixth node lists 'various XWiki extensions', none of which relate to notification settings or management.

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 following pages, notification settings, or 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: 0.53

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

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 0.25 because the relevant node, which is the fourth node, is ranked lower than three irrelevant nodes. The first node discusses 'domain-based authentication and session cookie configuration,' which is unrelated to the query about enabling user accounts upon first login with OpenID Connect. The second node provides 'general information about OpenID Connect authentication' but lacks specific details on automatically activating user accounts. The third node covers 'compatibility and installation details' of the OpenID Connect authenticator, but not the necessary configuration for automatic activation. The relevant node, which mentions the `xwiki.properties` file, should be ranked higher as it directly addresses the configuration needed for automatic user account activation.

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 details on adding or updating specific properties, restarting the XWiki instance, and the automatic activation of new user accounts upon first login via OpenID Connect.

CustomContextualRelevancy: The retrieval context does not contain information about updating the xwiki.properties file or the specific property oidc.enableUser for automatic user activation.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.43

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.75 because the answer partially addresses the question by discussing related topics, but it includes irrelevant information such as general sources and documentation links that do not directly answer the specific query 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 0.00 because all nodes in the retrieval context are irrelevant to the input query. The first node discusses 'the configuration of OpenID Connect endpoints' without mentioning user profile field mapping. The second node covers 'different authentication methods in XWiki' but not the mapping of OpenID Connect user profile fields. The third node provides 'general information about OpenID Connect authentication' without addressing user profile field mapping. The fourth node details 'OpenID Connect Authenticator extension' specifics without discussing user profile field mapping. The fifth node lists 'dependencies for the OpenID Connect Authenticator extension', which is unrelated to mapping user profile fields.

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention 'xwiki.properties' and 'oidc.user.mapping', which relate to customizing user profile fields (supporting sentences 1 and 2), they lack the example mapping code and specific instructions found in sentences 3 and 4 of the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information about customizing user profile fields in xwiki.properties or using oidc.user.mapping.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.3904761904761905

Individual_scores:

AnswerRelevancy: 0.9523809523809523

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant and provides useful information about setting up group synchronization between XWiki and an OpenID Connect provider. However, it slightly misses the mark by mentioning the lack of specific search results, which doesn't directly address the setup 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 0.00 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 or specific properties like 'oidc.groups.claim'. The second node covers various authentication methods, including OpenID Connect, but lacks specific information about group synchronization or configuration in 'xwiki.properties'. The third node mentions OpenID Connect authentication and provides links to configuration guides but does not specifically address group synchronization or the configuration steps needed in 'xwiki.properties'. The fourth node lists dependencies for the OpenID Connect authenticator extension but does not provide any information relevant to group synchronization or specific configuration properties. Lastly, the fifth node provides general information about the OpenID Connect authenticator, including installation and compatibility, but does not discuss group synchronization or the necessary configuration steps.

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 configuring group synchronization in the xwiki.properties file or related properties like oidc.groups.claim, oidc.groups.separator, or oidc.groups.mapping.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.64

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0.325

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.88 because the output mostly addresses the input query about configuring file upload restrictions, but it includes irrelevant links to pages about exporting features, XWiki syntax, and document lifecycle, which do not directly relate to the task of setting upload limits for images or PDFs.

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 and PDF export features', and 'various features and macros' that do not relate to attachment upload restrictions. The relevant nodes, which mention 'Size Limit', 'Mimetype Restriction', and provide 'detailed steps on setting the maximum size of an attachment', are ranked fourth and fifth. This lowers the score as the relevant information is not prioritized higher.

ContextualRecall: The score is 0.50 because while the 5th node in the retrieval context supports the steps for configuring the maximum upload size (related to sentences in Schritt 2), there is no information in the retrieval context about configuring MIME types or the 10 MB limit (related to sentences in Schritt 1).

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 10 MB.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.3777777777777777

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.89 because the response 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 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'validation patterns' which does not address script rights or administrative permissions. The second node talks about 'template providers and page creation', which is unrelated to the topic of script rights. The third node is about 'configuring a custom right check class', which also does not mention script rights or administrative permissions in XWiki.

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

CustomContextualRelevancy: The retrieval context does not mention the inability to deny rights implied by the admin right in XWiki, which is the key information in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.33793103448275863

Individual_scores:

AnswerRelevancy: 0.6896551724137931

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.69 because the output contains several irrelevant statements, such as source links and example structures, which do not directly address the configuration of permissions in XWiki. However, the score is not lower because the relevant parts of the output likely provide some useful information on configuring permissions in XWiki.

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 do not relate to configuring access controls for team areas. The second node is about 'notifications and extensions in XWiki', which also does not provide information on setting access rights for team areas. Subsequent nodes continue to address unrelated topics such as 'nested pages and subwikis', 'managing users and groups', and 'group synchronization and customization', none of which directly address 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 related to access control steps, group creation, or rights management as outlined in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.4875

Individual_scores:

AnswerRelevancy: 0.9375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant to the issue of the notification bell not being available, with only a minor irrelevant mention of XWiki syntax documentation.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on ensuring the actual output is perfectly aligned 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 does not relate to enabling notifications in the wiki settings. The second node lists 'various notification-related options and links', but fails to address enabling notifications through the `xwiki.properties` file. The third node

describes 'different notification preferences and macros', yet does not mention enabling notifications in the `xwiki.properties` file. The fourth node provides a 'description of the notifications application and its features', but does not relate to enabling notifications in the wiki settings. Lastly, the fifth node discusses 'RSS feeds and notification settings', but does not mention enabling notifications through the `xwiki.properties` file. All nodes should be ranked lower as they do not address the input query.

ContextualRecall: The score is 0.50 because the first sentence of the expected output is supported by the 5th node in the retrieval context, which discusses displaying notifications about events on the wiki. However, the second sentence about setting 'notifications.enabled' in 'xwiki.properties' lacks direct support from any node in the retrieval context.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.5609523809523809

Individual_scores:

AnswerRelevancy: 0.9047619047619048

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.90 because the response mostly addresses the issue of not receiving notifications, but includes irrelevant links that do not directly 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node 'lists links and topics related to notifications' without providing specific instructions on enabling notifications, which is not useful for the input. The second node 'provides a list of features and settings related to notifications' but lacks direct guidance on enabling notifications, making it irrelevant. The third node 'discusses RSS feeds and notification settings' but does not address enabling notifications directly, which does not help the input. The fourth node 'lists various notification preferences macros' without providing actionable steps, which is not relevant to the input.

ContextualRecall: The score is 0.50 because while nodes in the retrieval context support the general idea of enabling notifications and following users or pages (as seen in sentences 1, 2, 3, and 4 of the expected output), they lack specific details such as configuring 'xwiki.properties', the exact process of enabling alerts, and viewing followed users in the 'network' tab, which are mentioned in sentences 2, 4, 5, and 6.

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.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8492307692307692

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses the question about enabling notifications for personal 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. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context is ranked first, and it 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.' Great job on getting it right!

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 retrieval context.

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 well with the expected output. However, it lacks specific mention of 'Options de filtrage avancées' in the user profile settings.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.9292307692307693

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses how to disable notifications, but includes irrelevant 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 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant node. The first three nodes effectively address the issue of 'automatic page following and notification settings', which is directly related to the input query about disabling automatic notifications. The irrelevant node, ranked fourth, discusses 'RSS feeds and general notification settings', which does not pertain to the specific issue at hand.

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 includes information about auto-watched pages and custom notification filters, aligning with the expected answer. However, it lacks explicit mention of modifying settings for 'Suivi automatique des pages' and removing filters in 'Filtres personnalisés'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.42823529411764705

Individual_scores:

AnswerRelevancy: 0.9411764705882353

Faithfulness: 0.8

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

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 Authenticator, which slightly detracts from the overall focus.

Faithfulness: The score is 0.80 because the actual output incorrectly claims that a server restart is necessary after modifying 'xwiki.cfg', whereas the retrieval context specifies that no restart is required when using the LDAP Application.

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 settings', which do not relate to enabling user accounts automatically upon first login using OpenID Connect. The second node mentions 'container authentication and LDAP', lacking information about automatic user activation with OpenID Connect in XWiki. The third node, despite mentioning OpenID Connect, focuses on 'general authentication and synchronization' without specific instructions for automatic activation. The fourth node lists 'dependencies for the OpenID Connect extension' but does not address the necessary configuration for automatic user activation. Lastly, the fifth node is about 'LDAP and Active Directory integration', which is unrelated to the query.

ContextualRecall: The score is 0.40 because while nodes in the retrieval context mention 'OpenId Connect Authentication' and the 'xwiki.properties file', they lack specific details about adding or updating properties, saving the file, restarting XWiki, and automatically activating new user accounts on first login via OpenID Connect, which are crucial elements in sentences 1, 3, and 5 of the expected output.

CustomContextualRelevancy: The retrieval context does not contain information about adjusting the xwiki.properties file or the specific property oidc.enableUser for activating user accounts on first login.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.38095238095238093

Individual_scores:

AnswerRelevancy: 0.9047619047619048

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.90 because the response is mostly relevant and addresses the question about customizing user profile fields mapping in XWiki. However, it includes a source link that does not directly contribute to the solution, 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. The first node discusses 'authentication management and OpenID Connect provider endpoints,' but does not mention mapping user profile fields. The second node provides information about 'container authentication and other authentication methods,' which does not address mapping user profile fields. The third node describes 'the OpenID Connect Authenticator extension, its developer, and installation details,' but lacks information on mapping user profile fields. The fourth node lists 'dependencies for the OpenID Connect Authenticator extension,' which is unrelated to mapping user profile fields. The fifth node provides 'configuration guides for specific systems and mentions configuration files,' but does not detail how to map user profile fields.

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 lacks information about customizing user profile fields using 'oidc.user.mapping' in 'xwiki.properties', which is crucial in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5478205128205128

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.23076923076923078

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes source links that do not directly contribute to the explanation, 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 explicitly mentions 'group synchronization using the 'oidc.groups.mapping' property', is ranked third. The first node discusses 'general OpenID Connect provider configuration and endpoints', which is not related to group synchronization, and should be ranked lower. Similarly, the second node focuses on 'container authentication and Active Directory', which is also irrelevant to the input query. These irrelevant nodes being ranked higher than the relevant one results in a lower score.

ContextualRecall: The score is 0.23 because while the 3rd node in the retrieval context supports configuring group synchronization and mentions relevant properties like 'oidc.groups.claim' and 'oidc.groups.mapping', it lacks information on locating and editing the 'xwiki.properties' file, examples for property settings, and instructions on saving changes and restarting XWiki.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using 'oidc.groups.mapping', but lacks detailed steps on configuring 'oidc.groups.claim', 'oidc.groups.separator', and saving changes as outlined in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.36153846153846153

Individual_scores:

AnswerRelevancy: 0.8076923076923077

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.81 because the main content of the answer is relevant and addresses the question about limiting file uploads to images or PDFs of a specific size. However, the repeated irrelevant statements about source links that do not directly address the question prevent the score 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 question about configuring file upload restrictions. The first node discusses 'language preferences and translations', the second node talks about 'various applications and features', the third node is about 'PDF export customization', the fourth node lists 'features and macros', and the fifth and sixth nodes contain 'links to image files and references'. None of these nodes address the topic of file upload restrictions, leading to a score of 0.00.

ContextualRecall: The score is 0.00 because the retrieval context does not contain any relevant information or nodes that support the expected output, which includes details on file type and size restrictions, MIME type configurations, or specific instructions related to XWikiPreferences.

CustomContextualRelevancy: The retrieval context does not contain any information about configuring MIME types or setting maximum attachment size, which are key elements in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.7213725490196079

Individual_scores:

AnswerRelevancy: 0.8235294117647058

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.82 because the answer mostly addresses the question about refusing script rights to a space administrator. However, the inclusion of source links, which do not directly answer the question, slightly 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 0.58 because the first node in the retrieval context is about 'language preferences and translations in XWiki, which is unrelated to script rights or administrator permissions,' and should be ranked lower. Additionally, the fourth node discusses 'checking access rights programmatically, which does not directly address the question of refusing script rights to an administrator,' and the fifth node is about 'executing tasks with user rights and restoring contexts, which does not relate to the question of denying script rights to an administrator.' However, the second and third nodes are relevant as they discuss 'programming rights' and 'various rights, including ADMIN and SCRIPT,' which are pertinent to the input question.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the information in the retrieval context, specifically regarding XWiki's handling of administrator rights. 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, which is the core of the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.3619047619047619

Individual_scores:

AnswerRelevancy: 0.8095238095238095

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.81 because the response mostly addresses the question about configuring access rights in XWiki, but it includes several irrelevant links that do not directly relate to the specific task of setting view and edit 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 of a page in XWiki, but it does not provide any information about configuring access rights or spaces. The second node talks about modifying resources using XML and REST API, which is unrelated to setting access rights for spaces in XWiki. The third node is about managing comments and attachments, which does not relate to configuring space access rights in XWiki. The fourth node explains the concept of spaces and sub-pages in XWiki, but it does not address access rights or how to configure them. The fifth node is about migration actions and adapting code, which is not related to configuring access rights for spaces in XWiki. Lastly, the sixth node discusses configuration display settings in XWiki, which does not pertain to setting access rights for teams in 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. The retrieval context lacks information on specific steps for achieving access control in XWiki, such as creating groups, defining rights at different levels, or ensuring visibility and modification rights for teams.

CustomContextualRelevancy: The retrieval context lacks information on creating groups, setting rights at wiki and space levels, and inheriting rights for child pages, which are key points in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7087394957983193

Individual_scores:

AnswerRelevancy: 0.5294117647058824

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.53 because while the output contains some relevant information about authenticating users with access tokens, it is overshadowed by multiple irrelevant statements and fragments that do not directly address 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 0.50 because relevant nodes are not consistently ranked higher than irrelevant nodes. The first node in the retrieval context is about 'LDAP configuration and does not mention token-based authentication or JWT', which should be ranked lower. Similarly, the third node discusses 'OpenID Connect scopes and response types', which are unrelated to the expected output and should be lower. However, the second node is correctly ranked as it describes 'Authorized Applications Configuration', relevant to authenticating using tokens. The fourth node also supports the expected output by explaining 'how the extension allows external applications to authenticate users with JWT'.

ContextualRecall: The score is 0.71 because while the retrieval context provides detailed information about authenticating users with JWTs, including enabling the token authenticator and configuring authorized applications (supported by nodes 2 and 4 in the retrieval context), it lacks explicit details on generating JWT tokens with required claims and including the token in the 'Authorization' header, which are crucial steps mentioned in the expected output.

CustomContextualRelevancy: The retrieval context includes details on enabling the token authenticator, configuring authorized applications, and JWT requirements, but lacks specific information on JWT claims and the use of Ed25519 key.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.3625

Individual_scores:

AnswerRelevancy: 0.8125

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.81 because the response mostly addresses the question about authenticating users with access tokens, but it includes irrelevant source links and mentions the lack of information in search results, which 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 'model validation strategies in XWiki,' which is unrelated to user authentication using access tokens. The second node covers 'attachment validation and configuration,' which does not pertain to token-based user authentication. Lastly, the third node provides information on 'the admin UI for model validation,' which is not relevant to the question 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, as the content discussed is not covered by the nodes.

CustomContextualRelevancy: The retrieval context lacks any information related to token-based authentication or JWT, which are central to the expected answer.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.37333333333333335

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.87 because the response is mostly relevant to 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 access tokens'. The second node talks about 'user management features in XWiki' but does not address 'authentication with access tokens'. The third node covers 'rights management and user profiles', yet it lacks any mention of 'token-based authentication or JWTs'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by the nodes in the retrieval context, as the context does not mention token-based authentication, JWT functionality, steps for using the authenticator, or its limitations.

CustomContextualRelevancy: The retrieval context lacks any information about token-based authentication, JWT, or related configurations mentioned 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: 0.38

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.90 because the response is mostly relevant and addresses the issue about the 'Bell' icon in XWiki. However, it includes source links 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 is about 'installation instructions and prerequisites', which does not relate to enabling notifications. The second node describes the 'Alert menu and its functionalities' but lacks information on enabling notifications. The third node lists 'dependencies for the extension', which does not address the issue of enabling notifications. Lastly, the fourth node discusses 'overriding default notification templates', which is unrelated to the problem of enabling notifications.

ContextualRecall: The score is 0.00 because the expected output about enabling notifications in `xwiki.properties` is not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any mention of enabling notifications via the `notifications.enabled` setting in `xwiki.properties`, which is the core information in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.75

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

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 0.75 because the first and fourth nodes in the retrieval context are relevant, providing information on how to follow a user to receive notifications and listing ways to receive notifications in XWiki. However, the second node discusses filters and settings without addressing enabling notifications, the third node talks about notification RSS feeds without ensuring notifications are received, and the fifth node is about the 'What's New' application, which is unrelated to general notifications. These irrelevant nodes should be ranked lower than the relevant ones.

ContextualRecall: The score is 0.50 because while some details about enabling notifications and following users are supported by nodes in the retrieval context, key instructions like enabling notifications in `xwiki.properties` and ensuring subscriptions to pages or users are not directly covered.

CustomContextualRelevancy: The retrieval context covers following users and default watch settings but lacks specific instructions on enabling notifications in xwiki.properties and watching pages or wikis.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.96

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response directly and completely addresses the question about enabling notifications for one's own actions in XWiki without any irrelevant information. Great job!

Faithfulness: The score is 0.80 because the actual output incorrectly states there is an option to enable notifications for one's own actions, while the retrieval context clarifies that users do not receive such notifications by default unless the event specifically targets them.

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are 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,' which directly addresses the expected output. The second node mentions, '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. The irrelevant nodes, such as the third node about administrators configuring notifications, the fourth node discussing RSS feeds, and the fifth node providing an overview of different notification methods, do not address enabling notifications for one's own actions and are correctly ranked lower. Great job on achieving a perfect score!

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

CustomContextualRelevancy: The retrieval context fully contains the expected answer's information about default notification settings and the ability to disable the filter in the 'Advanced filtering options' of the user profile.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.7244444444444444

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 0.7555555555555555

ContextualRecall: 0.75

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.92 because the response effectively addresses how to disable notifications in XWiki, but includes an unnecessary source link that does not directly pertain to the question.

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

ContextualPrecision: The score is 0.76 because relevant nodes in the retrieval context, such as the first node mentioning 'settings' of the notifications and the third node discussing 'Advanced filtering options' and 'notification settings', are ranked higher. However, the score is not higher because some irrelevant nodes, like the second node discussing 'following a user and default watch settings' and the fourth node about 'enabling or disabling the email feature', are ranked above other relevant nodes.

ContextualRecall: The score is 0.75 because while some sentences in the expected output are supported by nodes in the retrieval context, such as the default behavior of page watching and notification settings (supported by nodes 2 and 1), the statement about being notified about changes on your work lacks direct support from any node.

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

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_005_result.json

Overall_score: 0.8366666666666667

Individual_scores:

AnswerRelevancy: 0.95

Faithfulness: 0.8333333333333334

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant and provides useful information on configuring OpenID Connect in XWiki. However, it includes a mention of LDAP authentication, which is not pertinent to the OpenID Connect setup, slightly affecting the score.

Faithfulness: The score is 0.83 because the actual output incorrectly claims that adjustments are needed in the xwiki.properties file for automatic account enabling upon first login using OpenID Connect, whereas the retrieval context states that user accounts are enabled on first login by default in the OpenID Connect configuration.

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context, which mentions the commented-out property '# oidc.enableUser=true', is ranked first. This is directly relevant to enabling user accounts on first login as mentioned in the expected output. Great job on getting the most relevant information right at the top!

ContextualRecall: The score is 0.40 because while the 1st node in the retrieval context supports the idea of enabling user accounts on first login, it lacks explicit mentions of adding or updating properties, saving the file, restarting the XWiki instance, and the automatic enabling of new user accounts via OpenID Connect.

CustomContextualRelevancy: The retrieval context contains the exact information about the 'oidc.enableUser' property being set to true, which matches the expected output.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_006_result.json

Overall_score: 0.95

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 1.0

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', is ranked first. This directly addresses the input query about customizing user profile fields,

ensuring that irrelevant nodes discussing access tokens, endpoints, group synchronization, and in-wiki configuration are ranked lower.

ContextualRecall: The score is 0.75 because the retrieval context nodes provide detailed information on mapping custom properties and using the `oidc.user.mapping` property (related to sentences 1-3 in the expected output), but they do not mention the need to save the `xwiki.properties` file and restart XWiki, which is covered in sentence 4.

CustomContextualRelevancy: The retrieval context contains all necessary information about customizing user profile fields using the `oidc.user.mapping` property in the `xwiki.properties` file, including examples and instructions to save and restart XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8611965811965814

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.46153846153846156

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant and provides useful information about configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes a source URL that doesn't directly address the configuration process, 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 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`, `oidc.groups.separator`, and `oidc.groups.mapping` properties, which are directly relevant to setting up group synchronization. The irrelevant nodes, ranked fourth and fifth, discuss installation instructions and authentication management, which do not directly address the group synchronization setup process. Great job on the ranking!

ContextualRecall: The score is 0.46 because while nodes in retrieval context provide specific examples and mention key properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', they lack the comprehensive step-by-step instructions present in the expected output, leading to a lower score.

CustomContextualRelevancy: The retrieval context contains detailed information on configuring group synchronization, including `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, matching the expected output. However, it lacks explicit instructions to save and restart the XWiki instance.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6206060606060606

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.6363636363636364

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the task of limiting uploads to images or PDFs of a certain size, but includes irrelevant source links that do not directly contribute to solving the 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 0.33 because the relevant node in the retrieval context, which provides information about setting the maximum size of attachments and mentions the configuration parameter in XWikiPreferences, is ranked third. However, irrelevant nodes such as the first node, which discusses linking to external files and images without addressing upload restrictions, and the second node, which provides general information about XWiki pages without covering attachment restrictions, are ranked higher.

ContextualRecall: The score is 0.64 because while nodes in the retrieval context provide detailed steps for configuring maximum upload size and mention the xwiki.properties file, they do not explicitly cover restricting uploads to only images or PDFs of maximum 10MB or configuring allowed mimetypes to block all other types.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions, but lacks specific details on configuring xwiki.properties or setting a 10MB limit.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.76

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

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 irrelevant source links that do not directly contribute to answering 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 0.00 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 admin rights. The second node lists 'various rights and their properties' without mentioning the denial of admin rights. The third node explains the 'Script' right without relating to denying admin rights. The fourth node is about 'checking access rights programmatically,' which does not address denying admin rights. Lastly, the fifth node discusses 'converting rights during page migration,' which is unrelated to denying admin rights.

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 imply other rights and

are not deniable.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7412745098039215

Individual_scores:

AnswerRelevancy: 0.7647058823529411

Faithfulness: 1.0

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.875

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.76 because the response addresses the main question about configuring permissions in XWiki, but includes multiple irrelevant source links that do not directly assist with the task at hand.

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, discussing 'rights inheritance and administration features' pertinent to configuring permissions in XWiki. However, the second node, ranked higher than it should be, focuses on 'parent/child relationship and nested pages concept,' which is not directly related to the query. Similarly, the third node discusses 'string formatting and OpenID Connect properties,' which are unrelated to configuring access rights. These irrelevant nodes being ranked higher than some relevant nodes, like the fourth node that 'explains the permission system in XWiki,' impacts the precision score.

ContextualRecall: The score is 0.88 because most sentences in the expected output are well-supported by the nodes in the retrieval context, particularly nodes 1, 4, and 6, which cover key aspects like access control, group creation, and rights setting. However, there is a lack of information 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, but lacks specific details about creating groups, setting wiki-wide and space-level rights, and ensuring only team members can edit their own space.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.4666666666666666

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0.1666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the issue of the unavailable bell icon, but includes several irrelevant statements that do not directly 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 0.17 because the relevant node in retrieval context, ranked 6th, discusses how administrators can configure notification settings and edit the xwiki.properties file, which is pertinent to enabling notifications. However, the first five nodes in retrieval context are irrelevant as they discuss topics like 'custom notification filters and their behavior', 'the alert menu and notification display', 'various ways to receive notifications', 'links to various notification-related features', and 'notification email settings and customization', none of which address enabling notifications in the wiki settings. This misranking of irrelevant nodes higher than the relevant one results in the low score.

ContextualRecall: The score is 0.50 because while the 5th node in the retrieval context supports the part about editing the xwiki.properties file, there is no explicit mention in any node about enabling notifications or setting `notifications.enabled` to `true`.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.391304347826087

Individual_scores:

AnswerRelevancy: 0.9565217391304348

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant to the issue of not receiving notifications, but it includes a source link that does not pertain to the problem, 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. The first node contains 'image links and references, which do not provide any information related to enabling notifications or following users and pages.' The second node is 'a collection of image links and references, unrelated to the topic of notifications or user settings.' The third node is 'a list of image links and references, offering no relevant information about notification settings or following users.' 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 the nodes in the retrieval context, which only contain image links and references.

CustomContextualRelevancy: The retrieval context contains image links and references, which do not provide any information related to enabling notifications or following users/pages as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.35

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.75 because the response addresses the main question about enabling notifications for one's own actions, but the inclusion of source links and irrelevant statements about the origin of information slightly detracts from its direct 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 and internationalization', the second node is about 'administrating wiki features', and the third node addresses 'importing documents and syntax highlighting'. None of these nodes relate to notification settings for user activity, which is the focus of the input.

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, leading to a complete lack of supportive evidence.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user profile options, which are the main points in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.3565217391304348

Individual_scores:

AnswerRelevancy: 0.782608695652174

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.78 because the output partially addresses the question about disabling notifications, but includes several irrelevant links to documentation pages that do not pertain to the issue of managing notifications.

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 'attaching files and managing attachments', which is unrelated to 'managing notifications or notification settings'. Similarly, the second node is about 'moving and deleting attachments', which does not pertain to 'notification settings or disabling notifications'. The third node lists 'various macros and their functionalities', none of which relate to 'notification settings or managing notifications'. The fourth node is about 'administrating wiki features like users and rights', which is not related to 'notification settings or managing notifications'. The fifth node discusses 'controlling permissions and security modules', which does not relate to 'notification settings or managing notifications'. Lastly, the sixth node lists 'various XWiki extensions', none of which are related to 'managing or disabling 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 contain any information related to automatic following of pages or notification settings as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.48

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

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. Great job on maintaining 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 'domain-based authentication and session cookie configuration', which does not relate to enabling user accounts upon first login with OpenID Connect. The second node provides 'general information about OpenID Connect authentication and installation instructions', missing the specific configuration needed. The third node focuses on 'technical details and metadata of the OpenID Connect Authenticator extension', without addressing automatic user activation. Lastly, the fourth node includes 'configuration properties related to OpenID Connect endpoints' but omits the property `oidc.enableUser` or the process for activating user accounts automatically.

ContextualRecall: The score is 0.40 because while the 4th node in the retrieval context supports the mention of 'xwiki.properties' and the need to uncomment properties related to OpenID Connect, there is no specific mention of adding or updating properties, restarting the XWiki instance, or ensuring automatic activation of new user accounts upon first-time login.

CustomContextualRelevancy: The retrieval context lacks any reference to adjusting the xwiki.properties file or enabling the oidc.enableUser property for automatic user activation.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.4

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about customizing user profile fields from an OpenID Connect provider 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. The first node discusses 'authentication management and OpenID Connect endpoints' but does not mention

'customizing user profile fields or mapping them in `xwiki.properties`'. The second node talks about 'various authentication methods like Container Authentication and LDAP' without addressing 'the customization of user profile fields from an OpenID Connect provider'. The third node describes 'the OpenID Connect Authenticator extension and its installation' but lacks information on 'user profile field mapping or the `xwiki.properties` file'. The fourth node provides 'metadata about the OpenID Connect Authenticator extension', which is unrelated to 'customizing user profile fields'. Finally, the fifth node lists 'dependencies for the OpenID Connect Authenticator extension' but does not cover 'mapping user profile fields or using `xwiki.properties`'.

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.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.4

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

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.

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 mention group synchronization or specific properties like 'oidc.groups.claim' or 'oidc.groups.separator'. The second node mentions OpenID Connect authentication and related extensions but lacks information on group synchronization or the specific configuration steps needed for it. The third node provides links and descriptions for OpenID Connect authentication and configuration guides but does not cover group synchronization or the specific properties required for it. The fourth node lists dependencies for the OpenID Connect Authenticator extension but does not relate to group synchronization or the configuration properties needed for it. The fifth node describes the OpenID Connect Authenticator extension and its compatibility but does not address group synchronization or the specific configuration steps required.

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 or instructions related to setting up group synchronization between XWiki and an OpenID Connect provider.

CustomContextualRelevancy: The retrieval context lacks specific details on group synchronization configuration in xwiki.properties, such as oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping, which are present in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.597121212121212

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 0.7272727272727273

ContextualPrecision: 0.325

ContextualRecall: 0.8

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the question about configuring upload restrictions in a wiki, but includes irrelevant links about exporting features, XWiki syntax, and document lifecycle, which do not directly relate to the query.

Faithfulness: The score is 0.73 because the actual output mentions parameters like \$wgMaxUploadSize and \$wgFileExtensions, which are not present in the retrieval context. Instead, the retrieval context discusses configuration in XWikiPreferences without referencing these specific parameters.

ContextualPrecision: The score is 0.33 because the relevant nodes in retrieval context, such as the fourth node which includes sections on 'Size Limit' and 'Mimetype Restriction', and the fifth node providing detailed steps on setting the maximum attachment size, are ranked lower than irrelevant nodes. For instance, the first node discusses 'trusted domains and PDF export customization', which are unrelated to configuring attachment restrictions, and should be ranked lower.

ContextualRecall: The score is 0.80 because the 5th node in the retrieval context supports the steps for configuring the maximum attachment size to 10 MB, including repeating the process for each (sub)wiki, which aligns with sentences in the expected output. However, it lacks details on configuring MIME types and the specific MIME types allowed, which are essential parts of the expected output.

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 10MB.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.34285714285714286

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about denying 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'form validation and JavaScript patterns', which is unrelated to denying script rights in XWiki. The second node is about 'templates and page creation in XWiki', which does not address the issue of script rights. The third node talks about 'a custom right check class in XWiki', but it does not specifically address the inability to deny script rights to an admin. Therefore, none of the nodes provide relevant information, resulting in a score of 0.00.

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

CustomContextualRelevancy: The retrieval context does not contain any information about rights or admin rights in XWiki, which is the main focus of the expected answer.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.4194444444444445

Individual_scores:

AnswerRelevancy: 0.7222222222222222

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.375

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.72 because the output partially addresses the question about configuring access permissions in XWiki, but includes several irrelevant links that do not directly help with the specific task of setting permissions for team areas.

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 control for teams. The second node is about the Notifications Application in XWiki, which does not relate to setting access permissions for team areas. The third node explains using nested pages and subwikis, which does not address how to configure access rights for teams in XWiki. The fourth node provides a link to managing users and groups but does not directly explain how to set access rights for team areas. Lastly, the fifth node is about group synchronization and customization in XWiki, which does not directly relate to configuring access control for team areas.

ContextualRecall: The score is 0.38 because only some aspects of the expected output are supported by the nodes in the retrieval context. Specifically, sentence 1 is supported by node 4 regarding group management, sentence 3 is supported by node 3 concerning nested pages and rights inheritance, but sentences 2, 4, 5, 6, and the introductory and reminder statements lack direct support from 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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.3

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because the output contains numerous irrelevant statements that do not address the issue of the notification icon not being available, such as links to unrelated documentation and discussions about unrelated features.

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, as they do not address enabling notifications in the wiki settings. The first node discusses 'custom notification filters and their behavior', which is unrelated. The second node lists 'various notification-related topics and links' without providing the needed information. The third node mentions 'notification preferences and macros' but lacks details on enabling notifications. The fourth node describes 'the notification alert menu and its features', which is not pertinent. Lastly, the fifth node talks about 'RSS feeds and notification settings' without addressing the required topic.

ContextualRecall: The score is 0.50 because the first sentence about ensuring notifications are enabled is supported by the 5th node in the retrieval context, which discusses displaying notifications. However, the second sentence, which gives specific instructions about setting 'notifications.enabled' in 'xwiki.properties', is not supported by 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'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.8233333333333335

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.87 because the response is mostly relevant to the issue of not receiving notifications, but it includes a source link that is not directly related to troubleshooting this specific 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 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant ones. The first node discusses 'Enable and disable the extension' and 'Enabling notifications for all users by default', which is directly related to ensuring notifications are activated. The second node includes 'Follow a user', aligning with the expected output's instruction on following users to receive notifications. The irrelevant nodes, ranked third and fourth, discuss 'Get a notification RSS feed' and 'various notification preferences macros', which do not directly address enabling notifications or following users. Great job on the ranking!

ContextualRecall: The score is 0.75 because while there are several supportive connections between the expected output and nodes in the retrieval context, such as enabling notifications and following users or pages (related to nodes 1, 2, and 3), there are missing details like the specific mention of 'xwiki.properties' and the 'network' tab for viewing followed users, which are not found in the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks specific details about setting 'notifications.enabled' in 'xwiki.properties' and using the alert

menu for following pages.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8492307692307692

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses how to enable notifications for one's own activity, but includes irrelevant source links that do not directly relate to 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, ranked first, directly addresses the issue of not receiving notifications for one's own activity and how to change that setting. It states, 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter,' which perfectly aligns with the input query. Great job on getting it spot on!

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' lacks support from 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, the specific location 'Options de filtrage avancées' is not explicitly mentioned.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.8766666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.83 because the main response addresses the question about disabling notifications effectively, but it includes irrelevant source links that do not directly relate to 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 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 'Auto Watch', which directly relate to modifying notification settings. The irrelevant node, ranked fourth, discusses 'RSS feeds and general notification settings', which is not directly related to disabling notifications for auto-watched pages.

Great job on getting everything in the right order!

ContextualRecall: The score is 0.75 because most sentences in the expected output are supported by nodes in the retrieval context, such as the 1st node relating to auto-watched mode and the 2nd node relating to notification preferences. However, the sentence about being notified of changes lacks direct support from the retrieval context.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.5018181818181818

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.91 because the response is highly relevant to configuring OpenID Connect in XWiki, but it includes a mention of LDAP Authenticator, which is not pertinent 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 0.00 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 automatically using OpenID Connect. The second node mentions OpenID Connect but lacks information on 'automatically activating user accounts on first login'. The third node provides a general description of OpenID Connect without addressing the specific configuration needed. The fourth node lists 'dependencies for the OpenID Connect extension' without discussing user account activation. Lastly, the fifth node focuses on 'LDAP and Active Directory integration', which is unrelated to the required OpenID Connect configuration.

ContextualRecall: The score is 0.60 because while the 3rd node in the retrieval context supports the general idea of configuring OpenID Connect in XWiki, it lacks specific instructions on updating the 'xwiki.properties' file and does not mention the need to save the file and restart XWiki.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.6338461538461538
Individual_scores:
AnswerRelevancy: 0.7692307692307693
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 0.4
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.77 because the response addresses the main question about customizing user profile fields in XWiki, but includes irrelevant statements such as providing a source link and offering further assistance, which 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 all relevant nodes are ranked higher than irrelevant nodes. The first node in the retrieval context is relevant as it mentions 'xwiki.properties', which directly relates to mapping fields as specified in the input. The subsequent nodes discuss topics like 'container authentication and LDAP', 'general information about the OpenID Connect Authenticator', 'dependencies for the OpenID Connect Authenticator', and 'configuration guides for specific systems', none of which address user profile field mapping, ensuring they are correctly ranked lower.

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention customizing user profile fields and using 'oidc.user.mapping', they lack explicit details on example mapping formats, replacing placeholders, and instructions to save and restart XWiki, which are present in the expected output sentences.

CustomContextualRelevancy: The retrieval context does not contain any information about customizing user profile fields or using the 'oidc.user.mapping' property in 'xwiki.properties'.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6155555555555555
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 0.3333333333333333
ContextualRecall: 0.4444444444444444
CustomContextualRelevancy: 0.3
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 present. 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 discusses 'group synchronization, including the use of claims in the xwiki.properties file and mapping between XWiki groups and provider groups', is ranked third. This means that irrelevant nodes, such as the first node which 'discusses authentication management and OpenID Connect provider configuration, but does not specifically address group synchronization', are ranked higher than the relevant node. Similarly, the second node 'focuses on container authentication, Active Directory, and LDAP, without mentioning group synchronization', which should also be ranked lower than the relevant node.

ContextualRecall: The score is 0.44 because while the 3rd node in the retrieval context supports the general process of configuring group synchronization in XWiki, including properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', it lacks specific details such as locating the 'xwiki.properties' file, providing examples for claim names and separators, and instructions on saving 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 as specified in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3444444444444444

Individual_scores:

AnswerRelevancy: 0.7222222222222222

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.72 because the main content addresses the question about configuring file upload restrictions effectively, but the repeated irrelevant source links lower 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 is unrelated to file upload restrictions. Similarly, the second node covers 'various applications and features in XWiki, such as tour guides and mail configuration', which do not relate to file upload restrictions. The third node describes 'technical details about PDF export and styling', which are not relevant to setting file upload restrictions. The fourth node lists 'various features and macros in XWiki, such as PDF viewers and office document viewers', which do not address file upload restrictions. The fifth node provides 'links to image files and references', which are not related to configuring file upload restrictions. Lastly, the sixth node again lists 'image references and links', which do not pertain 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. The retrieval context lacks information on file size or type restrictions, MIME type configuration, and specific configuration steps mentioned in the expected output.

CustomContextualRelevancy: The retrieval context lacks any information about configuring MIME types or setting maximum attachment sizes, which are essential details in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.8054545454545454

Individual_scores:

AnswerRelevancy: 0.7272727272727273

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.73 because the response partially addresses the question about denying script rights to a space administrator, but includes irrelevant source links that do not directly relate to the specific query, 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 the second node in the retrieval context is relevant as it 'discusses programming rights and script execution, which is directly related to the topic of script rights for administrators.' However, the first node, which is ranked higher, is irrelevant as it is 'related to language preferences and translations in XWiki, which is not relevant to denying script rights to an administrator.' Additionally, other irrelevant nodes like the third, fourth, and fifth nodes are ranked lower, but not all relevant information is prioritized effectively.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the retrieval context, specifically by the 3rd node, ensuring complete alignment. Great job!

CustomContextualRelevancy: The expected answer states that admin rights cannot be denied, which aligns with the retrieval context indicating 'ADMIN' is 'not deniable'. However, the retrieval context does not explicitly mention the lack of support for this in XWiki, which slightly reduces the score.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.3466666666666667

Individual_scores:

AnswerRelevancy: 0.7333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.73 because the output provides relevant information about configuring spaces in XWiki, but it includes multiple source links that do not directly address the input 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 provides 'XML metadata about a page in XWiki but does not address access control or permissions setup', which is unrelated to configuring team space access. The second node describes 'how to modify resources using XML representations and HTTP methods', which is also unrelated. Subsequent nodes similarly fail to address the input query, discussing topics like 'editing comments and managing attachments', 'the concept of spaces and pages', 'migration issues', and 'configuration display settings', none of which pertain 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 any information about creating groups, setting rights, or inheriting rights in XWiki as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7695238095238096

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.83 because the main content effectively addresses the question about authenticating users with access tokens, but the inclusion of multiple source links that 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!

ContextualPrecision: The score is 0.50 because the relevant nodes in the retrieval context, such as the second node which 'discusses configuring authorized applications for generating tokens' and the fourth node which 'provides a description of the extension that allows authentication using JSON Web Tokens', are ranked lower than some irrelevant nodes. For example, the first node 'is about LDAP authentication settings in XWiki, which is unrelated to token-based authentication' and should be ranked lower. Similarly, the third node 'describes OpenID Connect scopes and response types, which are not directly related to the use of access tokens' and should also be ranked lower. However, the score is not lower because the relevant nodes are still present and contribute to addressing the input question.

ContextualRecall: The score is 0.71 because while several key aspects of the expected output, such as the use of JWT for authentication and configuration details, are supported by nodes in the retrieval context, some specific instructions like generating a JWT with specific claims and including it in the 'Authorization' header are not covered.

CustomContextualRelevancy: The retrieval context includes details about enabling the token authenticator, configuring authorized applications, and using JWTs, matching most of the expected answer. However, it lacks specific mention of claims like expiration time and audience.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.3142857142857143

Individual_scores:

AnswerRelevancy: 0.5714285714285714

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.57 because while the response partially addresses the question about authenticating users with access tokens, it includes irrelevant details such as source links and unrelated strategies, which detract 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 discusses 'model validation strategies in XWiki,' which is unrelated to 'user authentication with access tokens.' The second node covers 'attachment validation and configuration in XWiki,' which does not pertain to 'token-based user authentication.' Lastly, the third node provides details on '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 key elements in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.3142857142857143

Individual_scores:

AnswerRelevancy: 0.5714285714285714

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.57 because the response partially addresses the question about authenticating users with access tokens, but includes irrelevant information about documentation, extensions, and general help that do not directly relate 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 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 talks about 'user rights and registration options' without mentioning 'authentication with access tokens'. The third node is about 'managing user rights and profiles', which does not address 'authentication using access tokens'.

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, indicating a complete lack of relevant information.

CustomContextualRelevancy: The retrieval context lacks any information about token-based authentication or JWT, which are central to the expected answer.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_gemma2_9B_Q4

File: qa_001_result.json

Overall_score: 0.3764705882352941

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the response is mostly relevant and provides useful information about the issue with the bell icon in XWiki. However, it includes a source link that does not directly address 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 0.00 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. The second node describes 'the functionality and appearance of the notifications menu', but does not provide information on enabling notifications. The third node lists 'dependencies for the notifications extension', which is not relevant to enabling notifications. The fourth node discusses 'overriding notification templates', which is unrelated to enabling notifications. Therefore, all nodes should be ranked lower as they do not address the input query.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by any nodes in the retrieval context, as there is no mention of enabling notifications or the 'notifications.enabled' setting in 'xwiki.properties'.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.7944444444444444

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.8055555555555555

ContextualRecall: 0.6666666666666666

CustomContextualRelevancy: 0.5

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 a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.81 because most relevant nodes in the retrieval context are ranked higher than irrelevant ones. The first node provides guidance on setting up notifications correctly, and the third and fourth nodes discuss enabling notification types and different ways to receive notifications, which are all relevant to the input. However, the second node, ranked second, discusses filters and settings related to hidden pages and personal activity, which does not directly address the issue of not receiving notifications, slightly lowering the score.

ContextualRecall: The score is 0.67 because while the retrieval context covers how to watch pages, locations, and users (as seen in the 1st node), it does not mention enabling notifications via 'xwiki.properties' or explicitly subscribing to pages or users.

CustomContextualRelevancy: The retrieval context mentions following users and watching locations for notifications, but lacks details on enabling notifications via xwiki.properties and watching pages or entire wikis.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 1.0

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

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 all relevant nodes in the retrieval context are perfectly ranked higher than the irrelevant ones. Great job!

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 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 in the user profile.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.75

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.45

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and directly addresses the question about disabling notifications in XWiki. Great job!

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

ContextualPrecision: The score is 0.45 because relevant nodes in the retrieval context are ranked lower than some irrelevant nodes. The first node discusses 'how to get a notification RSS feed and manage notification settings', but it does not address disabling automatic notifications, which is the main concern. The third node focuses on 'filtering notifications' without explaining how to disable automatic notifications. The fourth node is about 'enabling or disabling email notifications through configuration files', which is not directly related to the input query. The sixth node talks about 'overriding notification templates and installing extensions', which is unrelated to disabling automatic notifications. However, the second and fifth nodes are relevant as they mention 'Default Watch Settings' and 'Choosing the default auto-watched mode', directly addressing the issue of managing and disabling

automatic notifications.

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 alignment between the information provided and the expected details. Great job!

CustomContextualRelevancy: The retrieval context mentions default watch settings and user profile settings but lacks details on automatic page watching and custom filters as described in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8233333333333333

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.92 because the output is mostly relevant and addresses the question about configuring OpenID Connect in XWiki. However, it includes a mention of LDAP Authenticator, which is not relevant to the specific OpenID Connect configuration issue, 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 in the retrieval context, which mentions the configuration property '# oidc.enableUser=true', is ranked first. This node is directly related to enabling user accounts upon first login using OpenID Connect. Great job in ranking the relevant information at the top!

ContextualRecall: The score is 0.40 because while the 1st node in the retrieval context supports the mention of enabling user accounts on first login and includes the property '# oidc.enableUser=true', it lacks information about adding or updating properties, saving the file, restarting the XWiki instance, and the automatic enabling of new user accounts via OpenID Connect.

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 retrieval context lacks explicit instructions to save the file and restart the XWiki instance, which are part of the expected answer.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.8866666666666667

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and effectively addresses the question about customizing user profile fields in XWiki. However, it includes a source link that is not directly relevant to the specific customization query, preventing a perfect 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 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 using 'oidc.user.mapping', is ranked first. This directly addresses the input query about mapping additional user profile fields from OpenID Connect to XWiki. Great job on getting it spot on!

ContextualRecall: The score is 0.50 because while the retrieval context supports the concept of mapping custom properties using 'oidc.user.mapping' as mentioned in sentence 2 and 3 of the expected output, it lacks information on replacing placeholders with actual property names and the need to save and restart XWiki as stated in sentences 4 and 5.

CustomContextualRelevancy: The retrieval context contains all the 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8800000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

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 the relevant nodes in the retrieval context are ranked higher than the irrelevant nodes. The first three nodes provide essential information on configuring 'oidc.groups.claim' and 'oidc.groups.mapping', which are crucial for group synchronization as required in the expected output. The irrelevant nodes, ranked fourth and fifth, only provide general information about OpenID Connect and XWiki or discuss authentication management and endpoints, which do not directly address the group synchronization process.

ContextualRecall: The score is 0.50 because while the instructions for configuring properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' are supported by nodes in the retrieval context, key steps such as locating the 'xwiki.properties' file, opening it with a text editor, and saving and restarting XWiki are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context includes details on configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', aligning closely with the expected output. However, it lacks explicit instructions on locating and saving the 'xwiki.properties' file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6466666666666667

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.6

CustomContextualRelevancy: 0.3

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 0.33 because the relevant node in the retrieval context, which provides information on setting the maximum size of an attachment in XWikiPreferences, is ranked third. However, irrelevant nodes, such as the first node discussing linking to external files and images without addressing mimetype restrictions or maximum upload size configurations, and the second node explaining the structure of XWiki pages without relating to configuring attachment settings, are ranked higher. These irrelevant nodes should be ranked lower to improve the score.

ContextualRecall: The score is 0.60 because while the steps for configuring the maximum attachment size (Sentences 5-10) are well-supported by the 3rd node in the retrieval context, the instructions for configuring allowed mimetypes (Sentences 1-4) lack direct support from any specific node.

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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.6799999999999999

Individual_scores:

AnswerRelevancy: 0.4

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.40 because the output includes multiple irrelevant source links that do not address the question about denying script rights, which 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 question. The first node discusses programming rights and script execution but does not address denying rights to a space administrator. The second node lists various rights and their properties but does not mention the inability to deny rights to a space administrator. The third node explains the script right and its default settings but does not relate to denying rights to a space administrator. The fourth node provides code snippets related to access checks but does not address the specific issue of denying rights to a space administrator. The fifth node discusses converting rights during page

migration but does not mention denying 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 expected answer states that rights implied by admin cannot be denied, which aligns with the retrieval context indicating that admin rights are not deniable.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7901515151515152

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.875

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.91 because the response effectively addresses the configuration question, 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. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.67 because relevant nodes in the retrieval context are ranked higher, such as the first node discussing 'rights inheritance and administration features in XWiki' and the fourth node providing information on 'XWiki's permission system'. However, the score is not higher because the second node, which focuses on 'parent/child relationship and Nested Pages', and the third node, discussing 'string formatting and OpenID Connect properties', are ranked higher than the relevant fourth node. Additionally, the fifth node, which 'explains the creation of spaces and pages', is ranked higher than the relevant sixth node explaining 'XWiki's ability to set wiki-wide and page-level rights'.

ContextualRecall: The score is 0.88 because most of the expected output aligns well with nodes in the retrieval context, particularly nodes 4 and 6, which cover the key steps and concepts of access control in XWiki. However, the retrieval context lacks information on adjusting rights for additional teams or spaces, which slightly affects the overall alignment.

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and inheritance, which aligns with the expected answer's steps on setting wiki-wide and space-level rights. However, it lacks specific details on creating groups for each team and setting rights for 'XWikiAllGroup' and individual team spaces.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.5190476190476191

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to the issue of the notification bell not being available, with only a minor irrelevant closing statement that doesn't detract significantly 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 0.17 because the relevant node in the retrieval context, which mentions editing the 'xwiki.properties' file, is ranked last (6th node). The earlier nodes are irrelevant as they discuss topics like 'custom notification filters' (1st node), 'alert menu and notification display' (2nd node), 'various ways to receive notifications' (3rd node), 'links to notification settings' (4th node), and 'notification email settings and auto-watch modes' (5th node), none of which address enabling notifications in the wiki settings. This misordering significantly lowers the precision score.

ContextualRecall: The score is 0.50 because while the retrieval context mentions editing the xwiki.properties file (node 5), it does not specify the 'notifications.enabled' setting or setting it to 'true', which are key details in the expected output.

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

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_012_result.json

Overall_score: 0.3333333333333333

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the issue of not receiving notifications, but includes an irrelevant statement about sources and documentation, which doesn't relate to the input.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant, as they contain 'image links and references that are unrelated to the topic of notifications or their settings.' This means there are no relevant nodes ranked higher than irrelevant ones.

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 aspect of the expected instructions, such as enabling notifications, following pages or users, and viewing followed users, is unsupported by the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_013_result.json

Overall_score: 0.3

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.50 because the output includes source links that do not directly address the question about enabling notifications for one's own actions, making the response 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 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 notifications.' The third node covers 'various applications and extensions like Office Importer and Syntax Highlighting, but does not address notification settings or user activity notifications.' This means that no relevant information was retrieved, resulting in the lowest possible score.

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.

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

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_014_result.json

Overall_score: 0.3411764705882353

Individual_scores:

AnswerRelevancy: 0.7058823529411765

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.71 because the main content addresses how to disable notifications, but the presence of multiple 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 query. The first node discusses 'attaching files', the second node is about 'moving and deleting attachments', and the third node covers 'macros and their functionalities', none of which relate to 'notification settings or disabling notifications'. Similarly, the fourth node is about 'administrating wiki features', the fifth node lists 'various extensions', and the sixth node mentions 'applications and extensions', all of which do not address the input query about 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 about automatic page following, notification settings, or notification filters as mentioned in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: en

File: qa_015_result.json

Overall_score: 0.475
Individual_scores:
AnswerRelevancy: 0.875
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.5
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses the question about configuring XWiki for OpenID Connect, but the incomplete 'Empfehlung:' statement 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 query. The first node discusses 'domain-based authentication and session cookie configuration', which does not relate to enabling user accounts upon first login with OpenID Connect. The second node offers a 'general description of OpenID Connect authentication' but lacks mention of the specific property `oidc.enableUser` needed for automatic user activation. The third node includes 'metadata about the OpenID Connect authenticator', which is not pertinent to the configuration change required. Lastly, the fourth node talks about 'various OpenID Connect endpoints and properties' without addressing the `oidc.enableUser` property or automatic activation of user accounts.

ContextualRecall: The score is 0.50 because while the 4th node in the retrieval context supports the mention of 'xwiki.properties' and OpenID Connect, it lacks specific details about adding or updating the 'oidc.enableUser' property and the automatic activation of new user accounts upon first login.

CustomContextualRelevancy: The retrieval context does not mention the property 'oidc.enableUser' in xwiki.properties or its configuration for automatic user account activation with OpenID Connect.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_016_result.json

Overall_score: 0.3333333333333333
Individual_scores:
AnswerRelevancy: 0.6666666666666666
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.67 because the answer partially addresses the question about customizing user profile fields in XWiki, but includes irrelevant information such as sources and documentation links that do not directly help with the customization 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect provider endpoints', the second node is about 'container authentication', the third node provides 'general information about OpenID Connect authentication and configuration guides', the fourth node describes 'the OpenID Connect authenticator extension and its features', and the fifth node lists 'dependencies for the OpenID Connect authenticator extension'. None of these contexts mention or address the mapping or customization of user profile fields, 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 the information in the nodes in retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.38

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.90 because the response is highly relevant to the input question about setting up group synchronization between XWiki and an OpenID Connect provider. However, it includes source links that do not directly address the specific setup process, which 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 authentication management and OpenID Connect endpoints without mentioning group synchronization or configuration related to 'oidc.groups.claim' or 'oidc.groups.mapping'. The second node provides information about container authentication and Active Directory, which is unrelated to the group synchronization process between XWiki and an OpenID Connect provider. The third node mentions automatic synchronization with an OpenID Connect provider but lacks specific details on configuring group synchronization, such as setting properties in 'xwiki.properties'. The fourth node lists dependencies for the OpenID Connect extension but does not provide any information about configuring group synchronization. Finally, the fifth node describes the OpenID Connect authenticator's general capabilities and compatibility but does not address group synchronization setup or configuration.

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 configuring group synchronization in the xwiki.properties file, which is the main focus of the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.5483333333333335
Individual_scores:
AnswerRelevancy: 0.8333333333333334
Faithfulness: 1.0
ContextualPrecision: 0.325
ContextualRecall: 0.0833333333333333
CustomContextualRelevancy: 0.5
Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the question about configuring upload restrictions, but includes some irrelevant 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 0.33 because the relevant nodes in the retrieval context, such as the fourth node mentioning 'Size Limit' and 'Mimetype Restriction,' and the fifth node providing detailed steps on setting 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 is about 'ODT and PDF export configurations,' and the third node contains 'a list of macros and features,' all of which are unrelated to configuring upload restrictions for attachments. These irrelevant nodes should be ranked lower to improve the score.

ContextualRecall: The score is 0.08 because only one sentence from the expected output (Step 2, Sentence 1) is supported by the nodes in the retrieval context, specifically node 5, while all other sentences lack support from the retrieval context.

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 size to 10 MB.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_019_result.json

Overall_score: 0.4
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
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 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 denying script rights in XWiki. The second node is about 'template providers and page creation', which does not relate to script rights or administrative permissions. 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. Hence, all nodes should be ranked lower as they do not match the input query.

ContextualRecall: The score is 0.00 because the expected output about XWiki not supporting certain rights is not found in any of the nodes in the retrieval context, which focus on validation, templates, and

custom right checks, but do not address the specific admin rights issue mentioned in the expected output.

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

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_020_result.json

Overall_score: 0.39631336405529954

Individual_scores:

AnswerRelevancy: 0.8387096774193549

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.14285714285714285

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.84 because the output mostly addresses the question about configuring team access and permissions in XWiki, but it includes several irrelevant links to documentation that do not directly relate to the specific configuration task.

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 'notifications and the XWiki Development Team', not about setting access rights for teams. The third node explains 'nested pages and subwikis', which does not directly address team-specific access rights. The fourth node provides a link to 'managing users and groups', but does not directly explain how to set specific access rights for teams. The fifth node discusses 'group synchronization and customization', which is not directly related to configuring access control for teams in XWiki.

ContextualRecall: The score is 0.14 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 lack direct support from any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to the steps for achieving access control mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_021_result.json

Overall_score: 0.4428571428571429

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the issue with the notification bell icon, but includes irrelevant information about XWiki syntax and a general offer for

further assistance, which 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 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 'custom notification filters and their behavior', which does not relate to enabling notifications in the wiki settings. The second node lists 'various notification-related features and settings' but fails to address enabling notifications through `xwiki.properties`. The third node describes 'different notification preferences and macros', without mentioning enabling notifications in the wiki settings. The fourth node provides a 'general description of the notifications application and its interface', again not addressing enabling notifications in the wiki settings. Lastly, the fifth node discusses 'RSS feeds and notification settings' but does not mention enabling notifications in the wiki settings.

ContextualRecall: The score is 0.50 because the first sentence is supported by the 1st node in the retrieval context, which discusses notification filters, but the second sentence is unsupported as there is no mention of setting the `notifications.enabled` property 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 are key details in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_022_result.json

Overall_score: 0.7736363636363637

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.82 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. Great job maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. Great job on the perfect ranking!

ContextualRecall: The score is 0.75 because most of the expected output is supported by nodes in the retrieval context, such as enabling notifications and following users or pages. However, specific details like setting 'notifications.enabled' and the 'network' tab in the user profile are missing from the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' to true, following specific pages, and using the alert menu.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_023_result.json

Overall_score: 0.8266666666666665

Individual_scores:

AnswerRelevancy: 0.3333333333333333

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

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 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 issue of not receiving notifications for one's own activity and how to change this setting. The irrelevant nodes, which discuss topics like RSS feeds, lists of links, and general notification preferences, are appropriately 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, indicating 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 option to change this setting, aligning with the expected output. However, it does not explicitly mention the 'Options de filtrage avancées' section.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_024_result.json

Overall_score: 0.8736363636363637

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

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 with 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 correctly ranked higher than the irrelevant node. The first three nodes discuss topics like 'Choosing the default auto-watched mode', 'Create global notification filters', 'Notifications Auto Watch Preferences', and 'Default Watch Settings', which are directly related to modifying notification settings as expected. The fourth node, which is less relevant as it focuses on 'RSS feeds and general notification settings', is appropriately ranked last. Great job on maintaining perfect precision!

ContextualRecall: The score is 0.75 because the supportive reasons show that most sentences in the expected output are related to the 3rd node in the retrieval context, which discusses 'auto watched pages' and user preferences. However, the unsupportive reason indicates that the retrieval context does not cover notifications about changes to one's work, affecting the score.

CustomContextualRelevancy: The retrieval context includes information on 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.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_025_result.json

Overall_score: 0.44666666666666666
Individual_scores:
AnswerRelevancy: 0.8333333333333334
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.4
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.83 because the response is mostly relevant to configuring OpenID Connect in XWiki, but it includes an irrelevant statement about LDAP authentication, which is not needed for the specific 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. The first node discusses 'domain-based authentication and session cookies,' which is unrelated to the automatic activation of user accounts with OpenID Connect. The second node focuses on 'container authentication' and only mentions OpenID Connect in passing, without addressing the specific configuration needed. The third node provides general information about OpenID Connect but lacks details on the 'specific property `oidc.enableUser` needed for automatic user activation.' The fourth node lists dependencies for the OpenID Connect extension, which does not pertain to the configuration question. Lastly, the fifth node discusses 'LDAP and Active Directory integration,' which is unrelated to the input query.

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention 'OpenId Connect Authentication' and 'xwiki.properties', they lack details on adding or updating properties, saving the file, restarting XWiki, and guaranteeing automatic activation of new user accounts.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the xwiki.properties file or the specific property oidc.enableUser for automatically activating user accounts with OpenID Connect in XWiki.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_026_result.json

Overall_score: 0.41333333333333333
Individual_scores:
AnswerRelevancy: 0.6666666666666666
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.4
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.67 because the response partially addresses the question about customizing and mapping user profile fields in XWiki, but includes irrelevant information about sources 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. The first node discusses 'the authentication management class and OpenID Connect provider endpoints', which does not mention user profile field mapping or customization. The second node focuses on 'authentication methods like Container Authentication and Active Directory', without addressing user profile field mapping or customization. The third node provides 'general information about the OpenID Connect Authenticator', but does not mention mapping user profile fields or customizing them. The fourth node lists 'dependencies for the OpenID Connect Authenticator extension', which is unrelated to user profile field mapping or customization. The fifth node describes 'configuration guides and limitations for the OpenID Connect Authenticator', but does not address user profile field mapping or customization.

ContextualRecall: The score is 0.40 because while the retrieval context supports the general concept of customizing user profile fields and mapping them using 'oidc.user.mapping' (as seen in the 1st node in retrieval context), it lacks specific examples and details found in sentences 2, 3, and 4 of the expected output, such as the mapping example, replacing property names, and instructions to save and restart XWiki.

CustomContextualRelevancy: The retrieval context does not contain information about customizing user profile fields or mapping them in xwiki.properties, which is the main focus of the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6333333333333334

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.5

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.

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 retrieval context, which provides information on group synchronization using the 'oidc.groups.mapping' property and the 'xwiki.properties' file, is ranked third. However, irrelevant nodes are ranked higher, such as the first node which discusses general OpenID Connect provider configuration without addressing group synchronization, and the second node which talks about container authentication without providing details on group synchronization.

ContextualRecall: The score is 0.33 because while the 3rd node in the retrieval context supports configuring group synchronization, claims, and mapping, there is no support for specific actions like opening the file with a text editor, replacing placeholders, or saving and restarting XWiki.

CustomContextualRelevancy: The retrieval context mentions group synchronization using 'oidc.groups.mapping' but lacks details on 'oidc.groups.claim' and 'oidc.groups.separator'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3583333333333333

Individual_scores:

AnswerRelevancy: 0.7916666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.79 because the response mostly addresses the question about limiting file types and sizes in a wiki, but it includes multiple irrelevant source links that do not directly relate to the topic, 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. The first node discusses 'language preferences and translations in XWiki', which is unrelated to configuring file upload restrictions. The second node is about 'various XWiki applications and features, such as tour guides and administration', which do not relate to file upload settings. The third node focuses on 'PDF export customization using CSS and XSL', which does not pertain to file upload restrictions. The fourth node lists 'various macros and features in XWiki, such as PDF Viewer and Office Document Viewer', which are not relevant to configuring file upload limitations. The fifth and sixth nodes contain 'image links and references to various projects', which do not provide information on setting 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, which lacks information on file types, size limitations, MIME type configurations, and specific steps for setting attachment size limits.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.7233333333333334

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.83 because the main content effectively addresses the question about refusing script rights to a space administrator. However, the inclusion of an irrelevant source link repeated multiple times 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 0.58 because the first node in the retrieval context is about 'language preferences and translations in XWiki', which is unrelated to the question about script rights, 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', both of which do not directly address the question and should be ranked lower than the second and third nodes that are

relevant to script rights and permissions.

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 mentions that certain rights like ADMIN are not deniable, but it does not explicitly state that administrator rights cannot be refused as mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_030_result.json

Overall_score: 0.425

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.25

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the response largely addresses the question about configuring team spaces in XWiki, but it includes multiple irrelevant source links that do not directly assist with the configuration process.

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 discusses 'XML metadata about a page', which is unrelated to configuring access rights in XWiki. The second node talks about 'sending representations to modify resources', which does not pertain to setting access permissions for spaces. The third node focuses on 'editing and managing comments', which is not relevant to configuring team access. The fourth node explains 'the creation of spaces and sub-pages', but does not address access rights or permissions. The fifth node provides 'a list of actions related to page migration and renaming', which does not relate to access control settings. Finally, the sixth node discusses 'configuration display scopes', which does not provide information on setting access rights for teams.

ContextualRecall: The score is 0.25 because while there are connections between defining rights at the wiki and space levels to nodes in the retrieval context, key elements such as creating groups for teams and inheriting rights for child pages are not supported by any nodes in the retrieval context.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_031_result.json

Overall_score: 0.6695238095238094

Individual_scores:

AnswerRelevancy: 0.3333333333333333

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.33 because the actual output contains multiple irrelevant source links that do not address the question about authenticating users with access tokens, 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 0.50 because the second node, which is relevant as it 'discusses the configuration of 'Authorized Applications', which is relevant to using access tokens for authentication', is ranked after the first node, which is irrelevant as it 'is about LDAP configuration in XWiki, which is unrelated to token-based authentication or JWT'. Additionally, the fourth node, which 'provides a description of the extension that allows authentication with JWT, matching the expected output's description of token-based authentication', is ranked after the third node, which 'is about OpenID Connect scopes and response types, which is not directly related to JWT-based authentication as described in the expected output'. However, the relevant nodes are still present, contributing to the current score.

ContextualRecall: The score is 0.71 because while the retrieval context nodes provide substantial information on authenticating users with JWTs, enabling the token authenticator, and configuring authorized applications, they lack details on generating a JWT token with specific claims and signing key requirements, as well as including the token in the 'Authorization' header.

CustomContextualRelevancy: The retrieval context includes details about configuring authorized applications and enabling the token authenticator, which aligns with the expected answer. However, it lacks specific mention of JWT claims and the use of Ed25519 keys.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.27999999999999997

Individual_scores:

AnswerRelevancy: 0.4

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.40 because the output contains information about URLs and validation that are not directly related to the question about authenticating users with access tokens, which reduces 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 'Model Validation' and 'Name Strategies' in XWiki, which are unrelated to token-based authentication or JWTs. The second node is about 'Attachment Validation' and 'mimetype restrictions', not relevant to user authentication using access tokens. The third node provides information about the 'Admin UI for Model Validation' and its installation, which does not relate to the authentication process using 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 key concepts such as 'Benutzer mit Zugriffstoken', 'LLM-Anwendung (BETA)-Erweiterung', 'JSON Web Tokens (JWT)', and does not provide any related instructions or limitations as described in the expected output.

CustomContextualRelevancy: The retrieval context contains no information related to token-based authentication or JWT, which are central to the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_033_result.json

Overall_score: 0.3333333333333333

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

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 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. The first node discusses 'upgrading XWiki' and lacks any mention of 'authentication or token-based authentication'. The second node talks about 'user and rights management features' without addressing 'token-based authentication'. Lastly, the third node refers to an 'authentication guide' but fails to provide 'specific information about 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 is the focus of the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

Model: AI.Models.qa_llama3_1_402b

File: qa_001_result.json

Overall_score: 0.35666666666666663

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 0.9166666666666666

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.87 because the response is mostly relevant and addresses the issue of the missing 'Bell' icon in XWiki. However, it includes source links that do not directly address the problem, slightly reducing the score.

Faithfulness: The score is 0.92 because the actual output incorrectly claims that the text 'Installable with the Extension Manager' indicates if an extension can be installed, while the retrieval context states that offline installation is not supported in XWiki.

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 relate to enabling notifications in XWiki. The second node describes 'the functionality of the notifications menu and the bell icon' without providing solutions to enable notifications. The third node lists 'dependencies for the notifications extension,' which is unrelated to the issue at hand. Finally, the fourth node talks about 'customizing notification templates,' which does not address enabling notifications in the wiki settings. Thus, all nodes are irrelevant, leading to a score of 0.00.

ContextualRecall: The score is 0.00 because the retrieval context does not provide any information related to enabling notifications in the wiki or setting `notifications.enabled` in `xwiki.properties` to `true`, which are the key points in the expected output.

CustomContextualRelevancy: The retrieval context lacks any mention of the `notifications.enabled` setting or `xwiki.properties`, which are crucial for enabling notifications as described in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.6833333333333333

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.6666666666666666

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 1.00 because the output 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 states that notifications are not received until application types are enabled, whereas the retrieval context indicates that certain application types are enabled by default. Additionally, the actual output inaccurately claims that users don't receive notifications about their own activity by default, while the retrieval context specifies that this is true only for XWiki 14.1+ unless the event targets them explicitly.

ContextualPrecision: The score is 0.75 because the first and fourth nodes in the retrieval context are relevant, providing information on how to follow a user to receive notifications and outlining different ways to receive notifications in XWiki. However, the second node discusses filters and settings related to hidden pages and personal events, which does not directly address the issue of not receiving notifications and should be ranked lower. Similarly, the third node provides information about RSS feeds and notification settings but does not address enabling notifications or subscribing to specific pages or users, and the fifth node mentions the 'What's New' application, which is not directly related to setting up or troubleshooting notifications in XWiki. These irrelevant nodes are ranked higher than they should be, affecting the score.

ContextualRecall: The score is 0.50 because while the retrieval context supports some aspects like enabling notifications through the alert menu (sentence 4), following a user (sentence 5), and viewing watched users in the network tab (sentence 6), it lacks information on enabling notifications via `xwiki.properties` (sentence 2) and subscribing to pages or users (sentence 3).

CustomContextualRelevancy: The retrieval context includes information on following users and notification settings but lacks details on enabling notifications in xwiki.properties and watching pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.96

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

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

Faithfulness: The score is 0.80 because the actual output inaccurately describes the method for administrators to change users' notification settings, which is not detailed in the retrieval context. The retrieval context mentions the capability starting from XWiki 13.2 but lacks the specific method outlined in the claim.

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 the issue of enabling notifications for one's own actions by mentioning the ability to disable the filter in notification settings. Irrelevant nodes, such as the third node, which focuses on administrator configurations, are ranked lower, ensuring the precision score is perfect. Great job!

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 fully contains the expected answer, mentioning that by default one doesn't receive notifications about one's own activity unless explicitly targeted, and that this can be disabled in the 'Advanced filtering options' section 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: 0.61

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about disabling 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 0.00 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 it does not address 'disabling notifications for pages you are not interested in.' Similarly, the second node talks about 'following a user and default watch settings,' which also does not provide information on how to disable notifications for unwanted pages. Each subsequent node continues to miss the specific requirement of disabling notifications for pages you are not interested in.

ContextualRecall: The score is 0.75 because while nodes in the retrieval context support the information about default page watching settings (sentence 1), accessing notification settings (sentence 3), and managing custom filters (sentence 4), there is no support for being notified about changes on your work (sentence 2).

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8377777777777778

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant and provides useful information about configuring OpenID Connect in XWiki. However, it includes a mention of LDAP authentication, which is not pertinent to the question, slightly reducing the score.

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 node in the retrieval context, which includes the line '# oidc.enableUser=true' directly related to the configuration needed, is ranked first. This ensures that users receive the most pertinent information immediately, leading to a perfect score. Great job!

ContextualRecall: The score is 0.40 because while the retrieval context includes the key detail about enabling user accounts with the 'oidc.enableUser=true' property (as mentioned in sentences 1 and 2 of the expected output), it lacks information on the procedural steps such as adding or updating properties, saving the file, and restarting XWiki, which are crucial for the complete implementation described in sentences 3 to 5.

CustomContextualRelevancy: The retrieval context includes the property '# oidc.enableUser=true' which matches the expected answer about enabling user accounts on first login. However, it lacks explicit instructions on uncommenting and setting it to 'true', saving the file, and restarting XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.8733333333333334

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.87 because the answer mostly addresses how to customize user profile fields from an OpenID Connect provider in XWiki, but it includes a source link and a vague statement 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 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', is ranked first. This directly addresses the input question about customizing user profile fields. The irrelevant nodes, such as the second node focusing on 'scopes and response types for OpenID Connect' and the third node discussing 'endpoints and custom HTTP headers', are correctly ranked lower as they do not pertain to user profile field customization.

ContextualRecall: The score is 0.50 because while the retrieval context supports the mapping of custom properties using `oidc.user.mapping` as mentioned in sentences 2 and 3 of the expected output, it lacks details on replacing placeholders with actual values and the steps to save and restart XWiki as described in sentences 4 and 5.

CustomContextualRelevancy: The retrieval context contains detailed instructions on mapping user profile fields using the `oidc.user.mapping` property, including examples that match the expected output, and confirms the need to save and restart XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8800000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

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 the irrelevant nodes. The first three nodes provide crucial information on configuring the 'oidc.groups.claim' and 'oidc.groups.mapping' properties, which are essential for group synchronization as required in the input. The irrelevant nodes, ranked fourth and fifth, focus on installation instructions and authentication management, which are not directly related to the group synchronization process. Great job on the ranking!

ContextualRecall: The score is 0.50 because while nodes in the retrieval context provide specific details about properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', many sentences in the expected output are general instructions or introductory statements that do not directly match any specific part of the retrieval context.

CustomContextualRelevancy: The retrieval context includes details about configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', which align with the expected output. However, it lacks explicit instructions to save the file and restart the XWiki instance.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6766666666666665

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.4166666666666663

ContextualRecall: 0.6666666666666666

CustomContextualRelevancy: 0.3

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 a perfect alignment between the actual output and the retrieval context. Great job!

ContextualPrecision: The score is 0.42 because the relevant nodes in the retrieval context are not ranked as high as they should be. The first node discusses 'linking to external files and backlinks indexation,' which is not relevant to configuring upload restrictions or size limits, yet it is ranked higher than a relevant node. Similarly, the second node provides 'a general overview of XWiki's structure and features,' which does not address attachment upload restrictions, but is ranked above nodes that do. The third node, which explains 'how to set limits on the maximum size of an attachment in the XWikiPreferences document,' is relevant and should be ranked higher. Additionally, the fourth node, which discusses 'attachment display and mimetype restrictions,' is relevant and should also be ranked higher than the irrelevant nodes.

ContextualRecall: The score is 0.67 because while the 3rd node in the retrieval context provides detailed steps related to setting limits on attachment size and configuring mimetype restrictions, it does not explicitly mention configuring allowed mimetypes or the specific line 'attachment.upload.allowList=image/*,application/pdf'.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.7646153846153847

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.92 because the response is mostly relevant to the question about denying script rights to a space administrator. However, it includes a statement about content organization and nested pages migration, which is not directly related to the main topic, and an introductory example statement that doesn't directly address the 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 0.00 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 inability to deny script rights to a space administrator.' The second node lists 'various rights and their properties,' but does not mention 'denying script rights to a space administrator.' The third node provides 'details about script rights,' but does not state 'denying these rights to a space administrator is unsupported.' The fourth node is about 'checking access rights programmatically,' but does not relate to 'denying script rights to a space administrator.' The fifth node discusses 'converting rights during page migration' and does not address 'denying script rights to a space administrator.'

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

CustomContextualRelevancy: The expected answer states that rights implied by admin right cannot be denied, which aligns with the retrieval context indicating admin rights are not deniable. However, the retrieval context does not explicitly mention the lack of support for denying implied rights, slightly reducing completeness.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7412280701754386

Individual_scores:

AnswerRelevancy: 0.7894736842105263

Faithfulness: 0.75

ContextualPrecision: 0.6666666666666666

ContextualRecall: 1.0

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.79 because the output partially addresses the input by discussing permissions in XWiki, but includes several irrelevant links that do not directly help in configuring view and edit permissions as requested.

Faithfulness: The score is 0.75 because the actual output incorrectly claims that space-level rights override wiki-wide rights, while the retrieval context indicates that page-level permissions override wiki-wide permissions.

ContextualPrecision: The score is 0.67 because relevant nodes discussing 'rights inheritance and administration features' (first node), 'XWiki's permission system' (fourth node), and 'basic rules for setting rights in XWiki' (sixth node) are present. However, irrelevant nodes, such as the second node focusing on 'parent/child relationship and Nested Pages', the third node on 'string formatting and OpenID Connect user properties', and the fifth node on 'emulation of spaces and pages', are ranked higher than some relevant nodes, which affects the score.

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

CustomContextualRelevancy: The retrieval context discusses 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 wiki-wide rights, which are crucial steps in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.5333333333333333

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly 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 mentions 'editing the xwiki.properties file' to enable notifications, is ranked last (6th), while the irrelevant nodes, such as those discussing 'custom notification filters' (1st) and 'alert menu and notification display' (2nd), are ranked higher. This misranking of relevant information contributes to the low score.

ContextualRecall: The score is 0.50 because while the general topic of enabling notifications is supported by the 5th node in the retrieval context, there is no specific mention of the setting 'notifications.enabled' or the file 'xwiki.properties', which are key details in the expected output.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3904761904761905

Individual_scores:

AnswerRelevancy: 0.9523809523809523

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant to the issue of not receiving notifications, but it includes a source link that is unrelated to 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. The first node contains 'image links and references, which do not provide any information related to notifications or their settings.' Similarly, the second node is 'a list of image links and references, unrelated to the topic of notifications.' The third node also 'contains image links and references, which do not address the issue of enabling or managing notifications.' Therefore, none of the nodes are relevant to the input, 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.

CustomContextualRelevancy: The retrieval context contains only image links and does not provide any information related to enabling notifications or following users/pages as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.36

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.80 because the response is mostly relevant to the question about enabling notifications for one's own actions, but it includes a source link that does not directly address the specific issue, 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 '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 lists 'various XWiki extensions and applications, none of which relate to enabling notifications for user activity.' These irrelevant nodes are ranked higher than any relevant information, leading to a score of 0.00.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user profile options, which are central to the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.3375

Individual_scores:

AnswerRelevancy: 0.6875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.69 because the response partially addresses the question about disabling notifications, but it is lowered due to the repeated inclusion of irrelevant URLs that do not directly help in resolving the issue.

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 'attaching files and managing attachments', which is unrelated to 'notification settings'. The second node is about 'moving attachments', not about 'notification settings or disabling notifications'. The third node lists 'macros and their functionalities', which does not pertain to 'managing or disabling notifications'. The fourth node describes 'administration application and its

features', which are not related to 'notification settings'. The fifth node is about 'controlling permissions and security modules', which are not relevant to 'notification settings'. Finally, the sixth node lists 'various XWiki extensions', none of which relate to 'managing or disabling 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 about automatic page following, notification settings, or 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: 0.53

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about configuring XWiki for automatic user activation with OpenID Connect, 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 0.25 because the relevant node in the retrieval context, which discusses the `xwiki.properties` file necessary for enabling user accounts upon first login with OpenID Connect, is ranked fourth. The first three nodes in the retrieval context are irrelevant as they discuss 'domain-based authentication and session cookie configuration', 'general information about OpenID Connect and its installation', and 'OpenID Connect Authenticator compatibility without mentioning the `xwiki.properties` or `oidc.enableUser` property', which should be ranked lower.

ContextualRecall: The score is 0.40 because while the 4th node in the retrieval context mentions the xwiki.properties file and OpenID Connect, it does not provide specific instructions on adding or updating properties, saving the file, or restarting the XWiki instance, which are key details in the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information about modifying the xwiki.properties file or the oidc.enableUser property for automatic user activation.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.45999999999999996

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.90 because the response is highly relevant to the input question, providing useful information on mapping additional user profile fields from an OpenID Connect provider to XWiki user properties. However, the inclusion of 'Beispiel:' is slightly irrelevant as it doesn't add value to the explanation, preventing 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 OpenID Connect provider configuration', the second node is about 'container authentication and other authentication methods like LDAP and Active Directory', the third node provides a 'general description of the OpenID Connect Authenticator and its capabilities', the fourth node is about 'the OpenID Connect Authenticator's features and installation', and the fifth node lists 'dependencies for the OpenID Connect Authenticator extension'. None of these nodes address 'user profile field mapping or customization', which is the focus of the input.

ContextualRecall: The score is 0.40 because while the 1st node in the retrieval context supports the general idea of customizing user profile fields using 'xwiki.properties' and 'oidc.user.mapping', it lacks specific examples and instructions found in sentences 2 to 5 of the expected output.

CustomContextualRelevancy: The retrieval context lacks information about customizing user profile fields using 'xwiki.properties' and 'oidc.user.mapping', which are essential details in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.4

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the output 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 'various OpenID Connect endpoints and configuration properties' but lacks information on group synchronization. The second node talks about 'different authentication methods and extensions' without addressing group synchronization with OpenID Connect. The third node mentions 'OpenID Connect authentication and provides links to configuration guides' but does not cover group synchronization. The fourth node lists 'dependencies for the OpenID Connect authenticator' without information on group synchronization. Lastly, the fifth node provides 'general information about the OpenID Connect authenticator' but omits group synchronization details. All these nodes should be ranked lower as they do not address 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 retrieval context lacks any mention of key concepts such as 'Gruppensynchronisierung', 'xwiki.properties', or the specific configuration steps related to OpenID Connect provider integration with XWiki.

CustomContextualRelevancy: The retrieval context does not contain any information about configuring group synchronization in xwiki.properties, such as oidc.groups.claim, oidc.groups.separator, or oidc.groups.mapping.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6486363636363637

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0.325

ContextualRecall: 0.6

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the input regarding upload restrictions, but the inclusion of multiple source links that do not directly address the query 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 0.33 because the relevant nodes in the retrieval context are ranked lower than irrelevant nodes. The first node discusses 'trusted domains and PDF export customization,' which are 'not relevant to configuring attachment upload restrictions,' and should be ranked lower. The second node, about 'ODT export and PDF export support,' does not relate to 'setting MIME types or size limits for attachments,' and should also be ranked lower. The third node lists 'various macros and features unrelated to attachment upload settings,' which should be ranked lower than the relevant nodes. The fourth and fifth nodes, which include links to 'Size Limit' and 'Mimetype Restriction,' and provide 'detailed steps on setting the maximum size of an attachment,' are directly relevant and should be ranked higher.

ContextualRecall: The score is 0.60 because while the 5th node in the retrieval context provides detailed support for the steps related to configuring the maximum attachment size (sentences 8-14), it lacks information on configuring MIME types in `xwiki.properties` (sentences 2-7).

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and MIME type restrictions but lacks specific details about configuring xwiki.properties or setting 10 MB limit.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.36923076923076925

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.85 because the response is mostly relevant to the question about denying script rights to a space administrator, but it includes a source link that does not directly address the question, which slightly reduces 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 'JavaScript validation and programmatic validation', which is unrelated to denying script rights. The second node is about 'recommended templates and creating terminal pages', which does not pertain to rights management. The third node involves 'code for a custom right check', but it does not address 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 the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information about admin rights or their implications in XWiki, which is central to the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.35238095238095235

Individual_scores:

AnswerRelevancy: 0.7619047619047619

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.76 because the output addresses the main query about configuring team permissions in XWiki, but it includes several irrelevant source links that do not directly relate to the specific task of setting permissions, which prevents the score 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 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 in XWiki and does not relate to team access control configuration. The third node explains Nested Pages and subwikis in XWiki, which does not directly address the question of configuring team-specific access controls. The fourth node briefly mentions managing users and groups, but it does not provide specific information on setting access rights for teams as required by the input question. The fifth node is about group synchronization and customization in XWiki, which is not directly related to configuring team-specific access rights as asked in 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, as the retrieval context lacks any mention of steps for achieving access control, creating groups, setting rights, or inheriting permissions in XWiki.

CustomContextualRelevancy: The retrieval context lacks any information about setting access controls, creating groups, or managing rights in XWiki, which are key elements in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.4666666666666667
Individual_scores:
AnswerRelevancy: 0.8333333333333334
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.5
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.83 because the response is mostly relevant to the issue with the notification bell, but includes irrelevant 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'custom notification filters and their behavior', which does not relate to 'enabling notifications in the wiki settings'. Similarly, the second node provides 'links to various notification settings and customization options', but again does not mention 'enabling notifications in the wiki settings'. The third node lists 'various notification preferences and macros', which are not about 'enabling notifications in the wiki settings'. The fourth node describes 'the notification application and its features', but lacks information on 'enabling notifications in the wiki settings'. Lastly, the fifth node discusses 'RSS feeds and notification settings', which also does not address 'enabling notifications in the wiki settings'.

ContextualRecall: The score is 0.50 because while the 1st node in retrieval context supports the mention of 'notifications' and 'wiki' in sentence 1 of the expected output, there is no information in the nodes about 'xwiki.properties' or setting 'notifications.enabled' to 'true', which are crucial for sentence 2.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the xwiki.properties file or the specific property 'notifications.enabled'.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_022_result.json

Overall_score: 0.5236363636363637
Individual_scores:
AnswerRelevancy: 0.8181818181818182
Faithfulness: 0.9
ContextualPrecision: 0
ContextualRecall: 0.5
CustomContextualRelevancy: 0.4
Reasons:

AnswerRelevancy: The score is 0.82 because the response is mostly relevant to the issue of not receiving notifications, but it includes a source link that does not directly address the problem, which prevents the score from being higher.

Faithfulness: The score is 0.90 because the actual output incorrectly claims that notifications are enabled for all users by default, whereas the retrieval context specifies that only administrators have the ability to enable notifications for all users by default.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'various features and settings related to notifications' but does not address enabling notifications or following users. The second node mentions 'settings and filters' without providing information about enabling notifications or following users. The third node talks about 'RSS feeds and notification settings' but lacks details on enabling notifications in 'xwiki.properties' or following users. The fourth node describes 'various notification preferences macros' but fails to address enabling

notifications or following users.

ContextualRecall: The score is 0.50 because while some nodes in the retrieval context support the general idea of enabling notifications and following users or pages, specific details such as setting 'notifications.enabled' in 'xwiki.properties', using the 'alerte' menu, and viewing followed users in the 'réseau' tab are not covered.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' in 'xwiki.properties' and specifics on following pages or accessing user profiles.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.86

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.50 because the response includes irrelevant links that do not address the question about enabling notifications for one's own actions, but it still partially addresses the main concern.

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 is ranked first and 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,' which is precisely what the user is asking about. 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 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 ability to change this setting, aligning with the expected output. However, it lacks specific 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: 0.8766666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses how to disable notifications, but it includes irrelevant source 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 ranked higher than the irrelevant node. Great job on maintaining perfect precision!

ContextualRecall: The score is 0.75 because while the retrieval context nodes support most of the expected output, such as automatic page watching and notification preferences (related to nodes in retrieval context), it lacks information about being notified of changes on your work, which affects the overall alignment.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.39604395604395604

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 0.8571428571428571

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant to configuring OpenID Connect in XWiki, with only a minor irrelevant mention of LDAP authenticator that slightly detracts from the focus.

Faithfulness: The score is 0.86 because the actual output incorrectly claims that setting the session cookie on a top-level domain will encompass all wikis of the farm, while the retrieval context clarifies that this approach will not work if the domains are very different.

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 do not relate to enabling user accounts automatically with OpenID Connect. The second node talks about 'container authentication and LDAP', which also does not mention the specific configuration needed for OpenID Connect. The third node provides general information about OpenID Connect but omits the specific property `oidc.enableUser` required for automatic user activation. The fourth node lists dependencies for the OpenID Connect extension without addressing user account activation settings. Lastly, the fifth node is focused on 'LDAP and Active Directory integration', which is not relevant to the input query.

ContextualRecall: The score is 0.20 because only the first sentence of the expected output is partially supported by the information in the 1st node in retrieval context, which mentions OpenID Connect and xwiki.properties. The rest of the sentences, which provide specific instructions and guarantees, do not have corresponding details in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention adjusting the xwiki.properties file or setting oidc.enableUser to true for activating user accounts on first login.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.4049999999999997

Individual_scores:

AnswerRelevancy: 0.625

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.62 because the output contains some relevant information, such as providing sources and acknowledging the topic of customizing user profile fields in XWiki. However, it lacks direct and specific guidance on mapping additional user profile fields from an OpenID Connect provider to XWiki user properties, which 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 'authentication management and OpenID Connect endpoints' without mentioning user profile field customization. The second node is about 'different authentication methods and extensions, including OpenID Connect,' but lacks information on customizing user profile fields. The third node describes the 'OpenID Connect Authenticator extension' without addressing user profile field customization. The fourth node lists 'dependencies for the OpenID Connect Authenticator extension' without relating to user profile field customization. Finally, the fifth node provides 'a description and configuration guides for the OpenID Connect Authenticator,' but does not mention user profile field customization or mapping.

ContextualRecall: The score is 0.40 because while the nodes in retrieval context mention customizing user profile fields and the use of 'oidc.user.mapping', they do not explicitly cover the example mapping properties, the instructions for replacing property names and variables, or the steps to save and restart XWiki.

CustomContextualRelevancy: The retrieval context does not contain any information about customizing user profile fields or using the 'oidc.user.mapping' property in 'xwiki.properties'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5882051282051283

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.3076923076923077

CustomContextualRelevancy: 0.3

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.

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, directly addresses group synchronization by mentioning the use of claims and the 'oidc.groups.mapping' property, and is ranked after two irrelevant nodes. The first node discusses authentication management and endpoints without mentioning group synchronization, and the second node talks about authentication methods without providing details on group synchronization, both of

which should be ranked lower than the relevant node.

ContextualRecall: The score is 0.31 because while the 3rd node in the retrieval context provides some relevant information about configuring group synchronization in XWiki, it lacks details on specific steps such as locating and editing the 'xwiki.properties' file, setting properties with examples, and saving changes, which are crucial for a complete understanding.

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, which are present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3411764705882353

Individual_scores:

AnswerRelevancy: 0.7058823529411765

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.71 because the response mostly addresses the question about configuring file upload settings, but it includes multiple 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 query. 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 features, such as tour guides and mail configuration', which do not pertain to 'file upload settings'. The third node focuses on 'PDF export customization using CSS and XSL', which is not relevant to 'configuring file upload restrictions'. Subsequent nodes similarly fail to address the input query, as they cover topics like 'content viewing and notifications' and 'links to image files and references', none of which relate to 'file upload 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, as there is no mention of file size or type restrictions, MIME type configuration, or steps for setting maximum attachment size.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.7066666666666667

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.75 because the main content addresses the process of revoking script rights for a space admin, but it includes repetitive irrelevant source links that do not directly relate 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 0.58 because the first node in the retrieval context is irrelevant, discussing 'language preferences and translations in XWiki,' which should be ranked lower than the second node that directly addresses 'programming and script rights' relevant to the input. Additionally, the fourth and fifth nodes are also irrelevant, focusing on 'checking access rights programmatically' and 'executing code with user rights,' which do not pertain to denying script rights to an administrator. However, the second and third nodes are correctly identified as relevant, discussing 'programming rights' and 'ADMIN rights,' contributing to the current score.

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

CustomContextualRelevancy: The retrieval context mentions that certain rights like ADMIN are not deniable, but it does not explicitly state that rights implied by admin cannot be refused as in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.3555555555555555

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.78 because the main content addresses the question about configuring access rights in XWiki, 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 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 permissions for team spaces. The second node talks about 'sending representations to modify resources', which is not relevant to setting access permissions for team spaces in XWiki. Similarly, the third node is about 'managing comments and attachments on a page', which does not address configuring team space access permissions. The fourth node explains 'the concept of spaces and sub-pages in XWiki', but does not provide information on configuring access permissions for team spaces. The fifth node is about 'actions needed after running a migrator', which does not relate to configuring access permissions for team spaces. Lastly, the sixth node discusses 'the scope of configuration display in XWiki', which is not directly related to configuring 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 any information about creating groups, setting rights, or inheriting rights in XWiki, which are key elements in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7720879120879122

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.85 because the answer largely addresses the question about authenticating users with access tokens, but it includes multiple 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 first node discusses 'LDAP configuration' and does not mention token-based authentication, which is irrelevant to the input, yet it is ranked higher than the second node that explains 'Authorized Applications', relevant to using tokens for authentication.

Additionally, the third node focuses on 'OpenID Connect scopes', unrelated to token-based authentication, but is ranked higher than the fourth node that describes the use of JWT for authenticating users, directly supporting the input query.

ContextualRecall: The score is 0.71 because while several key aspects of the expected output are supported by nodes in the retrieval context, such as the use of JWTs and specific configuration steps, there are gaps in the context regarding the generation of JWT tokens with required claims and the inclusion of the token in the 'Authorization' header.

CustomContextualRelevancy: The retrieval context includes details about configuring authorized applications and enabling the token authenticator, which aligns with the expected answer. However, it lacks specific mention of JWT claims and the requirement to include the token in the 'Authorization' header.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.36

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.80 because the main content addresses the question about authenticating users with access tokens, but it includes unrelated source links and a discussion on document validation in XWiki, which detracts from the focus.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment 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 'model validation strategies and configuration in XWiki', which is unrelated to 'authentication using access tokens'. The second node is about 'attachment validation and configuration of mimetype restrictions', which does not pertain to 'user authentication with access tokens'. The third node provides information about 'the admin UI for model validation and its installation', 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 linked to any nodes in the retrieval context. The retrieval context lacks information on authenticating users with access tokens, JWT claims, Ed25519 keys, and the configuration or limitations of the authenticator.

CustomContextualRelevancy: The retrieval context lacks any information related to token-based authentication or JWT, which are central to the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.34285714285714286

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about authenticating users with access tokens, but includes unrelated URLs that do not 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 'upgrading and downgrading XWiki', which is unrelated to user authentication with access tokens. The second node provides information about 'user rights and registration options', but does not mention authentication with access tokens or JWTs. The third node focuses on 'managing user rights and profiles', without addressing token-based authentication or JWTs. Therefore, none of the nodes are relevant to the input query.

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 the retrieval context lacks information on token-based authentication for LLM applications, JWT usage, configuration steps, and authenticator limitations.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication, JWT, or the specific configurations mentioned 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: 0.335

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 0.8

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the response is mostly relevant and addresses the issue with the Bell icon in XWiki. However, it includes a source link that doesn't directly contribute to solving the problem, which slightly lowers the score.

Faithfulness: The score is 0.80 because the actual output incorrectly categorizes the Alert Notifications extension under 'Imported Extensions' instead of the 'Import category' as stated in the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node 'is about installation instructions for extensions and does not address enabling notifications in the wiki.' The second node 'describes the notifications menu and its features but does not provide information on enabling notifications.' The third node 'lists dependencies for the notifications extension but does not mention enabling notifications.' The fourth node 'discusses overriding notification templates but does not provide guidance on enabling notifications in the wiki.' These irrelevant nodes are ranked higher than any relevant ones, resulting in the lowest possible score.

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 'notifications.enabled' setting in 'xwiki.properties'.

CustomContextualRelevancy: The retrieval context lacks any mention of enabling notifications through the 'notifications.enabled' setting in 'xwiki.properties'.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.7759259259259259

Individual_scores:

AnswerRelevancy: 0.9629629629629629

Faithfulness: 1.0

ContextualPrecision: 0.75

ContextualRecall: 0.6666666666666666

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant to the question about not receiving notifications in XWiki, with only a minor irrelevant emotional expression that slightly detracts from a perfect score. Keep up the great work!

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 0.75 because the first node in the retrieval context is relevant, explaining how to ensure notifications are set up correctly by following a user. However, the second and third nodes are irrelevant, as they focus on filtering notifications and obtaining notifications through RSS feeds, which do not directly address the issue of not receiving notifications. The fourth node is relevant, providing an overview of notification methods in XWiki, which helps in understanding how to receive notifications. The presence of these irrelevant nodes ranked higher than some relevant ones is why the score is not higher.

ContextualRecall: The score is 0.67 because while several aspects of the expected output, such as watching pages, using the alert menu, following users, and checking the network tab, are supported by the nodes in the retrieval context, the crucial initial step of enabling notifications in `xwiki.properties` and the general concept of subscribing to pages or users are not covered in the 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 or wikis.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9866666666666667

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and provides useful information about enabling notifications for one's own actions in XWiki. However, the concluding remark 'That's it!' is slightly irrelevant, 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 all relevant nodes in the retrieval context are ranked higher than irrelevant ones, perfectly addressing the input question. Great job!

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 output information about default notification settings and the ability to disable the filter in the 'Advanced filtering options' section.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.7857142857142857

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 0.7

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is highly relevant and provides clear guidance on disabling notifications in XWiki. The only minor issue is the inclusion of a source link, which isn't directly necessary for answering the question.

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 0.70 because the first node in the retrieval context is relevant, discussing 'settings of the notifications' and 'enable or disable all notification types', which aligns with the input query about changing notification settings. However, the second node, which discusses 'following a user and default watch settings', and the third node, which is about 'filtering events and notifications', are not directly related to disabling notifications for pages and should be ranked lower. The fifth node is relevant as it mentions 'Choosing the default auto-watched mode', directly related to changing automatic page watching settings, but it is ranked lower than some irrelevant nodes.

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 contradictions or missing information. 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8904761904761905

Individual_scores:

AnswerRelevancy: 0.9523809523809523

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.95 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 OpenID Connect.

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 mentions the property '# oidc.enableUser=true' directly related to enabling user accounts on first login, is ranked first. All irrelevant nodes, discussing topics like domain-based authentication and cookie settings, general OpenID Connect information, and LDAP authentication, are ranked lower. Great job on getting it spot-on!

ContextualRecall: The score is 0.60 because while the retrieval context supports the configuration details in sentences 1 and 2 of the expected output through node 1 in the retrieval context, it lacks information on saving the file, restarting XWiki, and the automatic enabling of new user accounts as mentioned in sentences 4 and 5.

CustomContextualRelevancy: The retrieval context contains the key information about the 'oidc.enableUser' property, including its purpose and default setting, which aligns with the expected output. However, it lacks explicit instructions on saving the file and restarting XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.9214285714285715

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.86 because the output provides relevant information on customizing user profile fields in XWiki, but it includes several irrelevant statements that do not directly address the process of mapping additional fields from the OpenID Connect provider.

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 node in the retrieval context is relevant as it mentions 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user' and provides an example of how to map these properties, which directly relates to customizing user profile fields. This ensures that the relevant information is prioritized, resulting in a perfect score. Great job!

ContextualRecall: The score is 0.75 because the 3rd node in the retrieval context supports most of the expected output by discussing the customization of user profile fields and providing a similar example of mapping properties. However, it lacks information about saving the 'xwiki.properties' file and restarting XWiki, which affects the score.

CustomContextualRelevancy: The retrieval context includes the exact information from the expected output about mapping user profile fields using the `oidc.user.mapping` property in the `xwiki.properties` file, along with examples and instructions to save and restart XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8466666666666667

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.9

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 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first three nodes provide detailed information on configuring group synchronization, such as setting the `oidc.groups.claim`, `oidc.groups.mapping`, and `oidc.groups.separator` properties. The fourth node, ranked lower, is more about general information on authenticating with an OpenID Connect provider, and the fifth node discusses authentication management without specific details on group synchronization.

ContextualRecall: The score is 0.33 because only specific configuration details from the expected output, such as setting the 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' properties, are supported by nodes in the retrieval context. General instructions and introductory steps are not directly supported by the retrieval context.

CustomContextualRelevancy: The retrieval context includes configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', matching the expected answer. However, it lacks explicit instructions to save and restart XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6539393939393939

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.6363636363636364

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the output 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 0.33 because the relevant node in the retrieval context, ranked third, provides information on setting the maximum size of an attachment and mentions the configuration parameter in the XWikiPreferences document, which is directly related to configuring upload restrictions. However, the first and second nodes are irrelevant as they discuss linking to external files and images, and the structure of XWiki pages, which do not pertain to configuring upload restrictions for images or PDFs. Additionally, the fourth and fifth nodes are also irrelevant, focusing on attachment display, download, security settings, and versioning, which do not address configuring upload restrictions for specific file types or sizes.

ContextualRecall: The score is 0.64 because while the retrieval context nodes provide detailed steps for configuring maximum attachment size (node 3) and reference the configuration of mimetype restrictions (node 4), they do not directly cover the overall instruction to ensure only images or PDFs of max 10MB can be uploaded, nor do they include the specific line for mimetype restrictions.

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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.7428571428571429

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about denying script rights to a space administrator, but includes irrelevant information such as a link about

content organization and a statement that does not provide guidance.

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. For example, the first node discusses 'programming rights and script execution but does not address the inability to deny rights implied by admin rights.' Similarly, the second node lists 'various rights and their properties but does not mention the inability to deny rights that are implied by admin rights.' Each node is unrelated to the input query about denying script rights to a space administrator.

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

CustomContextualRelevancy: The expected answer states that rights implied by admin right cannot be denied, which is confirmed by the retrieval context indicating that admin rights are 'not deniable' and imply other rights such as LOGIN, VIEW, EDIT, DELETE, REGISTER, COMMENT, SCRIPT.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7547619047619047

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.75

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.86 because the answer mostly addresses the configuration question in XWiki, but it includes multiple irrelevant source links that do not contribute to solving the problem, 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 0.67 because the first node in the retrieval context is relevant, discussing 'rights inheritance and administration features' in XWiki, which aligns with setting view and edit permissions. However, the second and third nodes are irrelevant, focusing on 'parent/child relationship and Nested Pages' and 'string formatting and user identification in OpenID Connect', respectively, which do not address the configuration of permissions. These irrelevant nodes are ranked higher than the fourth and sixth nodes, which are relevant as they provide information on 'XWiki's permission system' and 'rules for setting rights in XWiki'.

ContextualRecall: The score is 0.75 because most of the expected output sentences are supported by nodes in the retrieval context, such as node 4 aligning with sentences 1, 2, and 3, and node 1 aligning with sentence 4. However, the summary sentences are not directly covered by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting rights at different levels and the concept of rights inheritance, 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 rights specifically for team spaces as outlined in the expected answer.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.5173333333333333

Individual_scores:

AnswerRelevancy: 0.92

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the response is largely relevant to the input, but it includes a source link that does not directly address the issue of the bell not being available, slightly lowering the 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 0.17 because the relevant node in the retrieval context, which is ranked sixth, mentions that administrators can edit the xwiki.properties file to enable notifications by setting 'notifications.enabled' to 'true'. However, the first five nodes are irrelevant as they discuss topics like custom notification filters, the Alert menu, different ways to receive notifications, and notification settings, but do not address enabling notifications or the 'notifications.enabled' setting. This misranking of irrelevant nodes higher than the relevant one results in a low score.

ContextualRecall: The score is 0.50 because while the retrieval context includes information about editing the xwiki.properties file (node 5), it does not explicitly 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'.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.38823529411764707

Individual_scores:

AnswerRelevancy: 0.9411764705882353

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant to the issue of not receiving notifications, with only a minor irrelevant detail about a source link that doesn't 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node contains 'image links and references to various projects and entities, which are unrelated to the topic of enabling notifications in XWiki.' The second node also includes 'image links and references to projects, which do not provide any information about notification settings in XWiki.' The third node similarly 'includes links and references to projects, without any relevant information about enabling notifications in XWiki.' These irrelevant nodes should be ranked lower than relevant ones.

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 only contain image links and references.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.37333333333333335

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.87 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.

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 retrieval context are irrelevant to the input query. The first node discusses 'customization of page templates, internationalization, accessibility, scalability, and document templates,' which are unrelated to notification settings. The second node provides 'information about an application for administrating wiki features,' which does not pertain to notification settings or user activity. The third node mentions 'various XWiki extensions like Office Importer, Font Awesome Icon Theme,' none of which relate to user notification settings or 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.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user profile options, which are essential parts of the expected answer.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.362962962962963

Individual_scores:

AnswerRelevancy: 0.8148148148148148

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.81 because the main content effectively addresses the question about disabling notifications, but it includes several irrelevant source links related to various XWiki topics that 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 to the input. For example, the first node discusses 'attaching files' which does not mention notifications or settings related to them. Similarly, the second node is about 'moving attachments', which is unrelated to notification settings. None of the nodes address the issue of managing or deactivating 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 does not contain any information related to automatic following of pages, notification settings, or 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: 0.40631578947368424

Individual_scores:

AnswerRelevancy: 0.631578947368421

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.63 because while the output contains some relevant information regarding OpenID Connect, it includes several irrelevant statements about setting `oidc.provider` and adjusting endpoints, which do not directly address the automatic activation of user accounts on first login.

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 with OpenID Connect. The second node provides 'general information about OpenID Connect authentication', but does not address the specific configuration needed. The third node offers 'metadata about the OpenID Connect authenticator', lacking any mention of automatic user activation. Lastly, the fourth node lists properties related to OpenID Connect, but omits the 'oidc.enableUser' property necessary for automatic activation.

ContextualRecall: The score is 0.40 because while some sentences in the expected output are supported by the 4th node in the retrieval context, which discusses 'xwiki.properties' and 'oidc.enableUser', key instructions such as updating the property, restarting the XWiki instance, and ensuring automatic activation of new user accounts are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about automatically activating user accounts on first login with OpenID Connect or modifying the xwiki.properties file.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.5062068965517241

Individual_scores:

AnswerRelevancy: 0.9310344827586207

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and provides a clear explanation on customizing user profile fields in XWiki. However, it includes source links that do not directly address the input question, slightly reducing 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 'authentication management and OpenID Connect endpoints' but lacks information on 'user profile field mapping or customization in xwiki.properties'. The second node is about 'container authentication and other authentication methods', not addressing 'mapping user profile fields from OpenID Connect to XWiki properties'. The third node provides 'configuration guides for OpenID Connect with various systems', without mentioning 'mapping user profile fields or customizing them in xwiki.properties'. The fourth node introduces the 'OpenID Connect Authenticator' but does not discuss 'user profile field mapping or customization in xwiki.properties'. Finally, the fifth node lists 'dependencies for the OpenID Connect Authenticator extension' but lacks details on 'user profile field mapping or customization in xwiki.properties'.

ContextualRecall: The score is 0.60 because while the retrieval context nodes provide information on customizing user profile fields and saving changes in `xwiki.properties`, they lack explicit details on example mappings and instructions for replacing placeholders, which are necessary to fully support the expected output sentences.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.4

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

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 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 'authentication management and OpenID Connect provider configuration', but not group synchronization. The second node focuses on 'authentication delegation' without mentioning group synchronization. The third node gives a 'general description of OpenID Connect authentication' but lacks details on group synchronization. The fourth node lists 'dependencies for the OpenID Connect extension', which is not relevant to group synchronization. Lastly, the fifth node

provides 'metadata about the OpenID Connect Authenticator extension', but does not address group synchronization.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by the nodes in the retrieval context, as all instructions and examples provided in the expected output are not found or detailed in the nodes.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6593675889328063

Individual_scores:

AnswerRelevancy: 0.8695652173913043

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.72727272727273

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.87 because the response mostly addresses the question about setting upload restrictions for attachments, but it includes irrelevant links to pages about exporting features, XWiki syntax, and document lifecycle features, which do not pertain to the specific 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 0.20 because the relevant node in the retrieval context, which provides 'detailed steps for setting the maximum size of an attachment and mentions mimetype restriction', is ranked fifth. Irrelevant nodes, such as the first node discussing 'trusted domains and PDF export look and feel', the second node about 'ODT and PDF export features', and the third node listing 'various macros and features related to notifications, PDF viewing, and other functionalities', are ranked higher than the relevant node.

ContextualRecall: The score is 0.73 because while the 5th node in the retrieval context provides comprehensive details and steps related to configuring the maximum attachment size, aligning well with sentences in the expected output, there is no information in the nodes about configuring MIME types or editing the xwiki.properties file, which are essential 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 to set the size to 10 MB.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.37142857142857144

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.86 because the response mostly addresses the input question about denying script rights to a domain 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. 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 'form validation and JavaScript patterns', which are unrelated to the question about script rights in XWiki. The second node is about 'templates and page creation in XWiki', which does not relate to script rights or administrative permissions. The third node involves 'code for a custom right check and configuration', but it does not address the specific issue of denying script rights to a space administrator in XWiki.

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

CustomContextualRelevancy: The retrieval context does not mention XWiki's admin rights or their limitations, which are central to the expected answer.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.37941176470588234

Individual_scores:

AnswerRelevancy: 0.6470588235294118

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.25

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.65 because the output partially addresses the question about configuring access and editing rights in XWiki, but includes several irrelevant statements related to other features and functionalities, which do not directly help in solving the specific problem presented.

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 use of relational databases and programming features, which is not relevant to configuring access control for teams. The second node is about notifications and application details in XWiki, which does not relate to setting access permissions for teams. The third node describes using nested pages and subwikis in XWiki, which is not directly related to configuring team-specific access rights. The fourth node mentions managing users and groups but lacks specific information on configuring access control for teams to view and edit areas. Lastly, the fifth node involves group synchronization and customization, which does not directly address the question of configuring team-specific access rights in XWiki.

ContextualRecall: The score is 0.25 because only sentences 1 and 2 in the expected output are supported by nodes in the retrieval context, specifically node 4, which mentions managing users and groups and setting user rights. The rest of the sentences, including area-level rights, inheriting rights, and summary statements, are not supported by the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information about creating groups, setting wiki-wide or area-specific rights, or inheriting rights for subpages, which are all key components of the expected answer.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_021_result.json

Overall_score: 0.4904761904761905
Individual_scores:
AnswerRelevancy: 0.9523809523809523
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.5
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant to the input, addressing the issue of the alert bell not being available. However, it includes a link to XWiki syntax documentation, which is not related to solving the problem, preventing a perfect score.

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', which is unrelated to the issue of 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', which does not help with the enabling issue. The fourth node describes 'the notification application and its features', but again, does not mention enabling notifications. Finally, the fifth node talks about 'RSS feeds and notification settings', which is not relevant to the problem of enabling notifications in the wiki settings.

ContextualRecall: The score is 0.50 because the first sentence can be linked to the 5th node in the retrieval context, which talks about displaying notifications on the wiki. However, the second sentence lacks support as there is no information in the retrieval context about setting 'notifications.enabled' in 'xwiki.properties' to 'true'.

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

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_022_result.json

Overall_score: 0.7388461538461538
Individual_scores:
AnswerRelevancy: 0.7692307692307693
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 0.625
CustomContextualRelevancy: 0.3
Reasons:

AnswerRelevancy: The score is 0.77 because the response mostly addresses the issue of not receiving notifications, but includes irrelevant source links and a generic note 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. Keep up the great work!

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node discusses 'Enable and disable the extension' and 'Enable and disable the email feature', directly relating to activating notifications. The second node covers 'Follow a user', aligning with instructions for receiving notifications. Irrelevant nodes, such as the third node discussing RSS feeds and the fourth node listing notification preferences macros without specific guidance, are ranked lower. Great job maintaining perfect precision!

ContextualRecall: The score is 0.62 because while several aspects of following users and pages for notifications are supported by nodes in the retrieval context, specific details like setting 'notifications.enabled' in 'xwiki.properties', using the 'alert' menu, and viewing followed users in the 'network' tab are not covered.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' to 'true', using the alert menu, and viewing followed users in the network tab.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8072727272727273

Individual_scores:

AnswerRelevancy: 0.6363636363636364

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.64 because the response partially addresses the question about enabling notifications for one's own actions, but includes several irrelevant statements that detract 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 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 default settings and how to change them. The irrelevant nodes, which discuss topics like RSS feeds, general notification settings, a table of contents, and various notification preferences macros, are all ranked lower, ensuring the most relevant information is prioritized. Great job on getting it spot on!

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 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, it does not explicitly mention 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.87363636363637
Individual_scores:
AnswerRelevancy: 0.81818181818182
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 0.75
CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.82 because the main content effectively addresses how to disable notifications, but the inclusion of source links that don't directly answer the question slightly detracts from the overall relevance.

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 the irrelevant node. The first three nodes provide information directly related to modifying notification settings and auto-watch preferences, which aligns perfectly with the input query. The irrelevant node, which discusses RSS feeds and notification types, is correctly ranked last, ensuring a perfect score. Great job!

ContextualRecall: The score is 0.75 because most of the expected output sentences are supported by nodes in the retrieval context, specifically regarding auto-watched mode and notification preferences in nodes 1 and 2. However, the sentence about being notified of changes lacks explicit support in the retrieval context, which affects the score.

CustomContextualRelevancy: The retrieval context includes information about auto-watched pages and custom filters, aligning with the expected answer's mention of automatic page following and custom filter settings. However, it lacks explicit details about modifying the 'Suivi automatique des pages' setting in user profile notification settings.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.38095238095238093
Individual_scores:
AnswerRelevancy: 0.9047619047619048
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.90 because the response effectively addresses the OpenID Connect configuration question, but includes irrelevant details about LDAP Authenticator and a general hope statement, 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. For instance, the first node discusses 'domain-based authentication and session cookie configuration,' which does not relate to 'enabling user accounts automatically upon first login using OpenID Connect.' Similarly, the second node mentions 'container authentication and other authentication methods like LDAP,' but fails to address the specific configuration for OpenID Connect. Additionally, the third node, despite discussing OpenID Connect, lacks 'specific instructions about modifying the `xwiki.properties` file to enable user accounts automatically.' 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 of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the xwiki.properties file or the specific property oidc.enableUser for activating user accounts via OpenID Connect.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.475

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the answer is mostly relevant and addresses the question about customizing user profile fields in XWiki using OpenID Connect. However, it includes source links that are not directly relevant to the customization process, which 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 the authentication management class and OpenID Connect endpoints, but not user profile field mapping or customization. The second node is about container authentication and other methods like LDAP and Active Directory, not OpenID Connect user profile field mapping. The third node provides general information about the OpenID Connect Authenticator extension without addressing 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. The fifth node provides links to configuration guides and examples for different systems, but does not discuss how to customize user profile fields in XWiki using OpenID Connect.

ContextualRecall: The score is 0.50 because while the retrieval context supports the customization of user profile fields and mapping in 'xwiki.properties' (as seen in the 1st node), it lacks specific details about replacing 'myxproperty1' and 'myxproperty2', and the need to save 'xwiki.properties' and restart XWiki, which are mentioned in the expected output.

CustomContextualRelevancy: The retrieval context lacks any mention of customizing user profile fields or mapping them in xwiki.properties, which is central to the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5766666666666667

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.25

CustomContextualRelevancy: 0.3

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.

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

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which discusses 'group synchronization and the use of the `oidc.groups.mapping` property', is ranked third. Irrelevant nodes, such as the first node that 'primarily discusses authentication management and endpoints configuration', and the second node that 'talks about container authentication and other authentication methods', are ranked higher than the relevant node. This misranking results in a lower score.

ContextualRecall: The score is 0.25 because while the 3rd node in retrieval context supports the general idea of configuring group synchronization in XWiki, it lacks specific details found in the expected output, such as locating the 'xwiki.properties' file, providing examples for 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', and instructions on saving and restarting XWiki.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using 'oidc.groups.mapping', but lacks details on 'oidc.groups.claim' and 'oidc.groups.separator'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.36875

Individual_scores:

AnswerRelevancy: 0.84375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.84 because the answer largely addresses the question about controlling file uploads on a wiki, but it includes multiple 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 'language preferences and translations in XWiki', which is unrelated to configuring file upload restrictions. The second node describes 'various XWiki applications and features, such as the Administration UI and PDF export', but does not address file upload settings. The third node focuses on 'technical details of PDF export customization', which is not relevant to setting file upload restrictions. The fourth node lists 'various XWiki macros and their functionalities', none of which pertain to file upload restrictions. The fifth node provides 'links to image files and references', which are not related to configuring file upload settings. The sixth node contains 'links to image files and references', which do not provide information on setting file upload restrictions.

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 the context lacks information on file types, size restrictions, MIME types, or configuration settings related to attachments.

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

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_029_result.json

Overall_score: 0.66
Individual_scores:
AnswerRelevancy: 0.6
Faithfulness: 1
ContextualPrecision: 0.5
ContextualRecall: 1.0
CustomContextualRelevancy: 0.2
Reasons:

AnswerRelevancy: The score is 0.60 because the response partially addresses the question about refusing script rights to a space administrator, but includes irrelevant information such as a source link about internationalization features and a statement indicating an inability to assist, which 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. Keep up the great work!

ContextualPrecision: The score is 0.50 because the second node in the retrieval context is relevant, discussing 'programming and script rights' and how 'script rights are allowed unless explicitly revoked', which aligns with the input query. However, the first node, which is ranked higher, is irrelevant as it talks about 'language preferences and translations in XWiki', not related to script rights or administrative permissions. Additionally, the third node, which is also irrelevant, 'lists various rights and their default states' without addressing the specific issue of refusing script rights to an administrator.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the information from the node in the retrieval context. 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 as in the expected output.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_030_result.json

Overall_score: 0.37037037037037035
Individual_scores:
AnswerRelevancy: 0.8518518518518519
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses the question about configuring spaces in XWiki, but it includes multiple source links that do not directly contribute to answering the specific configuration 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 provides 'XML metadata about a page, which is not relevant to configuring access

rights in XWiki.' The second node discusses 'sending representations and modifying resources using XML, which does not help in setting up access controls for spaces in XWiki.' The third node is about 'editing comments and managing attachments, which is unrelated to configuring team access in XWiki.' The fourth node explains 'the concept of spaces and sub-pages in XWiki, but does not provide information on setting access rights for teams.' The fifth node discusses 'migration actions and code changes, which are not relevant to configuring access rights for teams in XWiki.' Lastly, the sixth node describes 'configuration display settings in XWiki, which does not relate to setting access rights for teams.'

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 creating groups, defining rights at wiki or space level, or inheriting rights for child pages as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.8236507936507937

Individual_scores:

AnswerRelevancy: 0.8611111111111112

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.8571428571428571

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.86 because the response mostly addresses the question about authenticating users with access tokens, but includes multiple irrelevant source links and a non-technical comment that 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 0.50 because relevant nodes in the retrieval context are not consistently ranked higher than irrelevant nodes. The first node discusses 'LDAP authentication settings', which is unrelated to token-based authentication, yet it is ranked higher than the second node, which provides relevant information on 'configuring authorized applications' for JWTs. Similarly, the third node about 'OpenID Connect scopes' is not directly related to JWT-based authentication and is ranked higher than the fourth node, which directly supports the expected output by describing functionality for authenticating users with JWTs.

ContextualRecall: The score is 0.86 because most of the expected output is well-supported by the nodes in the retrieval context, particularly regarding the use of JWT tokens and configuration steps. However, the specific detail about including the token in the 'Authorization' header is not directly mentioned in any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context contains most of the expected answer details, such as enabling the token authenticator, configuring authorized applications, and the limitation on user creation. However, it lacks specific mention of JWT claims and the Ed25519 signing key.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.32727272727272727

Individual_scores:

AnswerRelevancy: 0.6363636363636364

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.64 because while the response partially addresses the question about authenticating users with access tokens, it includes multiple irrelevant statements about unrelated topics and source links, which detract 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 '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', which does not pertain to '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 retrieval 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 generieren', 'Authorization-Header', 'Bearer TOKEN', 'Erstellung oder Aktualisierung von Benutzern', or 'bestehenden Benutzer'.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.34285714285714286

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.71 because the response mostly addresses the question about authenticating users with access tokens, but it includes source links that are not directly relevant to the topic, which 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 'upgrading XWiki' which does not relate to 'authenticating users with access tokens'. The second node talks about 'features of the Administration Application' without addressing 'authenticating users with access tokens'. The third node covers 'managing user rights and profiles' but fails to mention '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.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication or JWT, which are key aspects of the expected answer.

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: 0.36470588235294116

Individual_scores:

AnswerRelevancy: 0.8235294117647058

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.82 because the response is mostly relevant to the question about the missing 'Bell' icon in XWiki, but it includes repeated irrelevant statements about source links and search results 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 0.00 because all nodes in the retrieval context are irrelevant to solving the issue of the bell not appearing. The first node discusses 'installation instructions and prerequisites for extensions,' which is unrelated. The second node describes 'the functionality and appearance of the notification bell and menu,' but does not help in enabling notifications. The third node lists 'dependencies for the notifications extension,' which does not address the problem. The fourth node explains 'how to override notification templates,' which is not relevant to the issue.

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 in `xwiki.properties` mentioned in the expected output.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.77

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.6

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 0.75 because the first and fourth nodes in the retrieval context are relevant to the input, discussing how to receive notifications and troubleshoot notification issues in XWiki. However, the second node, which is ranked higher than the fourth, discusses 'filtering notifications and hidden pages', and the third node focuses on 'RSS feeds and notification settings', neither of which directly address the issue of not receiving notifications. These irrelevant nodes should be ranked lower to improve the score.

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 process of enabling notifications via the 'alert' menu (node 4) and following a user (node 1), other key instructions like enabling notifications in `xwiki.properties` and ensuring subscriptions are not directly supported by any nodes.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.975

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.875

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

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

Faithfulness: The score is 0.88 because the actual output incorrectly claims that notifications for your own activities are typically enabled by default, while the retrieval context clarifies that users do not receive such notifications by default.

ContextualPrecision: The score is 1.00 because all relevant nodes in the retrieval context are ranked higher than irrelevant ones. The first two nodes directly address the issue by mentioning that 'By default, one don't receive notification about one's own activity' and 'This can be changed by switching off the Own event filter,' which are pertinent to enabling notifications for one's own actions. The irrelevant nodes, such as the third node discussing 'administrator settings and global notification configurations,' are ranked lower, ensuring high precision. 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 contains all information from the expected answer, 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 of the user's notification settings.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.7492063492063492

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8571428571428571

ContextualPrecision: 0.5888888888888889

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

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 0.86 because the actual output incorrectly mentions a 'Notification Level' setting to disable all notifications, which is not supported by the retrieval context. Instead, notifications are managed by enabling or disabling specific types or applications.

ContextualPrecision: The score is 0.59 because relevant nodes, such as the second node discussing 'Default Watch Settings' and the third node on filtering notifications, are ranked lower than some irrelevant nodes. The first node, which is about notification RSS feeds and settings, and the fourth node, focusing on email notifications at an administrative level, are not directly related to disabling notifications for specific pages and should be ranked lower. Additionally, the sixth node about overriding notification templates is also not relevant and should be ranked lower than the relevant nodes.

ContextualRecall: The score is 1.00 because all sentences in the expected output are fully supported by the nodes in the retrieval context, indicating a perfect match. 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.85

Individual_scores:

AnswerRelevancy: 0.95

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.95 because the output is highly relevant to configuring OpenID Connect in XWiki, with only a minor mention of LDAP Authenticator, which is not directly related 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 1.00 because the first node in the retrieval context is highly relevant, mentioning the property 'oidc.enableUser=true', which is directly related to enabling user accounts on first login. All irrelevant nodes, such as those discussing 'domain-based authentication and session cookies' (second node), 'general information about OpenID Connect' (third node), 'container authentication and LDAP' (fourth node), and 'LDAP configuration' (fifth node), are appropriately ranked lower. Great job on achieving perfect precision!

ContextualRecall: The score is 0.40 because while the retrieval context includes information about enabling user accounts on first login (related to sentence 1), it lacks details on adding or updating properties, saving the file, restarting the XWiki instance, and the automatic enabling of new user

accounts (related to sentences 2, 4, and 5).

CustomContextualRelevancy: The retrieval context contains the relevant property '#oidc.enableUser=true', matching the expected answer's instruction to uncomment and set it to true. However, it lacks explicit instructions to save the file and restart the XWiki instance.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.9433333333333334

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.8

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.92 because the response effectively addresses the question about customizing user profile fields in XWiki using OpenID Connect, but includes a source link that does not directly contribute to the solution, slightly affecting the 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 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 using 'oidc.user.mapping', is ranked first. This directly relates to customizing user profile fields, while the irrelevant nodes, such as the second node discussing scopes and response types, the third node focusing on endpoints and HTTP headers, and others, do not address user profile field mapping or customization and are correctly ranked lower.

ContextualRecall: The score is 0.80 because most of the expected output is supported by the 1st node in retrieval context, such as the customization of user profile fields and example mappings. However, the retrieval context does not mention saving the xwiki.properties file or restarting XWiki, which affects the score.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8682352941176472

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9411764705882353

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

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 0.94 because the actual output mentions the creation or update of specific XWiki groups ('xwiki-admins' and 'xwiki-users'), which are not specified in the retrieval context. This minor inconsistency slightly affects the faithfulness score.

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 on configuring the `oidc.groups.claim` and `oidc.groups.mapping` properties, which are crucial for synchronizing groups between XWiki and the OpenID Connect provider. The irrelevant nodes, ranked fourth and fifth, only provide general information about OpenID Connect authentication and configuration guides or discuss authentication management and endpoint configuration, which do not specifically address group synchronization.

ContextualRecall: The score is 0.50 because while nodes in the retrieval context provide specific details about configuring properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', several general and introductory instructions from the expected output are not explicitly covered in the retrieval context.

CustomContextualRelevancy: The retrieval context contains detailed information on configuring the `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` properties, matching the expected output. However, it lacks explicit instructions on locating and editing the `xwiki.properties` file and restarting the XWiki instance.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6166666666666668

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.8

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.75 because the output provides relevant information about restricting file uploads to images or PDFs of a certain size, but it includes irrelevant source links and suggestions about MediaWiki extensions that do not directly address 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 'describes how to set limits on the maximum size of an attachment', is ranked third. However, irrelevant nodes are ranked higher, such as the first node that 'discusses linking to external files and images and mentions upload summaries, but does not address configuring allowed mimetypes or setting maximum upload sizes', and the second node that 'provides general information about XWiki's structure and functionalities but does not relate to configuring upload restrictions or mimetypes'.

ContextualRecall: The score is 0.80 because most of the detailed steps for configuring mimetypes and setting maximum attachment size are well-supported by nodes in the retrieval context, specifically nodes 3 and 4. However, the introductory statement and the exact line for mimetype configuration 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 on configuring 'xwiki.properties' and setting the size to 10MB.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.77

Individual_scores:

AnswerRelevancy: 0.85

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.85 because the answer is mostly relevant and provides useful information, but it includes repeated irrelevant statements about a source link that does not directly address how to deny 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 script execution' but does not address denying script rights to a space administrator. The second node lists 'various rights and their properties' without mentioning the inability to deny rights implied by admin rights. The third node explains 'script right and its default settings' but lacks information on denying script rights to a space administrator. The fourth node contains 'code snippets related to access checks' but does not provide relevant information. The fifth node is about 'converting rights during page migration' and does not relate to the input question.

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 expected answer states that rights implied by admin right cannot be denied, which is supported by the retrieval context indicating that ADMIN right is not deniable and implies other rights.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.680952380952381

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 0.8

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.5714285714285714

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.87 because the response is mostly relevant to configuring access rights in XWiki, but it includes repeated irrelevant statements about source links that do not directly address the specific configuration question.

Faithfulness: The score is 0.80 because the actual output incorrectly claims that space creation can be restricted by setting 'Create Space' permission, whereas the retrieval context indicates that the concept of space has been removed from the XWiki UI.

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 detailing XWiki's

permission system, are ranked higher. However, the second node, which talks about 'parent/child relationship and nested pages,' and the third node, focusing on 'string manipulation and formatting,' are irrelevant and should be ranked lower than the relevant nodes. Additionally, the fifth node, which only discusses 'creation of spaces and pages,' is also irrelevant to the input question.

ContextualRecall: The score is 0.57 because while the retrieval context nodes provide information about XWiki's permission system and group management (nodes 4 and 6), they lack specific details on the outcomes and adjustments for additional teams or spaces, which are crucial for a complete understanding of the expected output.

CustomContextualRelevancy: The retrieval context mentions setting rights at wiki and space levels and inheriting rights, aligning partially with expected steps on setting view and edit rights and inheriting rights for child pages. However, it lacks specific details on creating groups for each team and setting rights specifically for XWikiAllGroup and team-specific groups.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.5

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the response is mostly relevant to the issue of the bell not being available, but it includes repeated irrelevant statements about providing a source link, which 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 0.17 because the relevant node in the retrieval context, which is the sixth node, mentions editing the `xwiki.properties` file to enable notifications by setting `notifications.enabled` to `true`. However, this relevant information is ranked lower than the first five nodes, which are irrelevant as they discuss topics like 'custom notification filters', 'notification bell display', and 'administrators' ability to configure settings' without addressing how to enable notifications.

ContextualRecall: The score is 0.50 because while the retrieval context (node 5) supports the general idea of configuring notifications through settings and editing the xwiki.properties file, it lacks specific mention of the 'notifications.enabled' setting and its value 'true' as described in sentence 2 of the expected output.

CustomContextualRelevancy: The retrieval context lacks information about enabling notifications via the xwiki.properties file, which is crucial in the expected answer.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.38571428571428573

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to the issue of not receiving notifications, but it includes an unrelated link, which slightly detracts from its 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. The first node contains 'a list of image links and references, which do not relate to enabling notifications or following users/pages in XWiki.' Similarly, the second node is 'a list of image links and references, which do not provide information on notification settings or following mechanisms in XWiki.' The third node also consists of 'image links and references, which are irrelevant to the topic of enabling notifications or following users/pages in XWiki.' These irrelevant nodes should be ranked lower than any relevant nodes, but since there are no relevant nodes, the score remains at 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 contains image links and references, but lacks any information about enabling notifications, following pages or users, or managing notification settings as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.34285714285714286

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.71 because the answer partially addresses the question about enabling notifications for one's own activities, but includes irrelevant statements like 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. The first node discusses 'themes, internationalization, accessibility, and performance,' which are 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 'importing office documents, icon themes, sharing pages, syntax highlighting, and mentions,' none of which relate to user 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 as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.3375

Individual_scores:

AnswerRelevancy: 0.6875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.69 because while the output addresses the question about disabling notifications, it includes several irrelevant links to documentation pages and extensions that do not directly relate to the user's query.

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 about topics like 'attaching files', 'moving and deleting attachments', 'macros and features for displaying content', 'administrating wiki features', 'permissions and security modules', and 'extensions and applications for XWiki', which are unrelated to notification settings or disabling notifications. None of the nodes address the input query about disabling 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 does not contain any information about automatic page following or notification settings as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.48

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

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 0.00 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 'automatically activating user accounts upon first login with OpenID Connect'. The second node provides 'general information about OpenID Connect authentication and configuration guides', lacking the specific details for automatic activation in XWiki. The third node contains 'metadata about the OpenID Connect Authenticator extension', which is unrelated to the needed configuration. Finally, the fourth node mentions the 'xwiki.properties file' but does not include the necessary property or instructions for enabling automatic user activation.

ContextualRecall: The score is 0.40 because while the 4th node in the retrieval context mentions 'xwiki.properties' and OpenID Connect, supporting the connection to the expected output about adjusting 'xwiki.properties', there is no specific mention of adding or updating properties, restarting

XWiki instances, or ensuring automatic activation of new user accounts, which are key aspects of the expected output.

CustomContextualRelevancy: The retrieval context does not contain information about adjusting the xwiki.properties file or the specific property oidc.enableUser for automatic user activation.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.6214285714285714

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.25

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.86 because the response is mostly relevant and provides useful information, but it includes statements that offer sources and documentation links without directly addressing the specific question about customizing user profile fields in XWiki.

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 highly relevant as it discusses the 'xwiki.properties' file and OpenID Connect, which directly pertains to configuring user profile fields. Great job!

ContextualRecall: The score is 0.25 because only the first sentence of the expected output is supported by the 1st node in the retrieval context, which discusses 'xwiki.properties' and OpenID Connect provider configuration. The rest of the expected output lacks support as the retrieval context does not mention 'oidc.user.mapping', provide examples, or instructions on applying changes.

CustomContextualRelevancy: The retrieval context lacks specific information about customizing user profile fields using 'oidc.user.mapping' in 'xwiki.properties' and the example mappings provided in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.36923076923076925

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.85 because the response provides relevant information and sources, which are helpful, but it doesn't directly address the specific process of setting up 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management and OpenID Connect endpoints' without mentioning group synchronization. The second node provides 'information on various authentication methods' but lacks details on group synchronization. The third node describes 'OpenID Connect Authenticator extension' capabilities without specific instructions for group synchronization. The fourth node lists 'dependencies for the OpenID Connect Authenticator' unrelated to group synchronization. Finally, the fifth node offers 'general information about the OpenID Connect Authenticator' without addressing group synchronization.

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 the context lacks any mention of group synchronization setup between XWiki and OpenID Connect or related configuration steps.

CustomContextualRelevancy: The retrieval context lacks details on configuring group synchronization properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6697058823529412

Individual_scores:

AnswerRelevancy: 0.8235294117647058

Faithfulness: 1.0

ContextualPrecision: 0.325

ContextualRecall: 0.7

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the question about setting upload restrictions for images or PDFs in a wiki, but includes irrelevant links about exporting features, 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 irrelevant, discussing 'trusted domains and PDF export customization', 'ODT and PDF export features', and 'various macros and features', none of which relate to configuring attachment restrictions in a wiki. These irrelevant nodes are ranked higher than the relevant nodes, which appear later and discuss 'Size Limit' and 'Mimetype Restriction', as well as detailed steps on setting maximum attachment size and MIME type restrictions. Additionally, the last node is also irrelevant as it focuses on 'document lifecycle features'.

ContextualRecall: The score is 0.70 because while the 5th node in the retrieval context provides detailed steps related to configuring the maximum attachment size (sentences 6 to 11 in the expected output), there is no information about configuring MIME types (sentences 2 to 5 in the expected output) in the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_019_result.json

Overall_score: 0.36363636363636365

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.82 because the main content effectively addresses the question about denying script rights to a space administrator, but the inclusion of source links that do not directly relate to 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 'form validation and JavaScript patterns', which are unrelated to denying script rights to a domain administrator. The second node is about 'template providers and page creation in XWiki', which does not relate to script rights or administrative permissions. The third node involves a 'code snippet for a custom right check class', but it does not address the specific issue of denying script rights to a domain administrator in XWiki.

ContextualRecall: The score is 0.00 because the expected output sentence about admin rights in XWiki is not supported by any information in the nodes in retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information related to admin rights or their denial, which is central to the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.37236842105263157

Individual_scores:

AnswerRelevancy: 0.7368421052631579

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.125

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.74 because the output partially addresses the question about configuring team access and permissions in XWiki, but includes several irrelevant source links that do not directly relate to the specific task of setting permissions for teams to view all areas while only editing their own.

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 do not relate to configuring access control for teams. The second node describes the 'Notifications Application in XWiki', which is unrelated to setting access permissions for team areas. The third node explains 'Nested Pages and subwikis', but lacks information on configuring access rights for teams. The fourth node mentions 'managing users and groups', but does not provide specific instructions on setting access rights for team areas. Finally, the fifth node discusses 'group synchronization and customization in XWiki', which is not directly related to configuring access control for team areas.

ContextualRecall: The score is 0.12 because only sentence 1 from the expected output is supported by node 4 in the retrieval context, which mentions managing users and groups. All other sentences lack support from the retrieval context nodes.

CustomContextualRelevancy: The retrieval context lacks any information related to access control steps or managing team-specific rights in XWiki, as outlined in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.4833333333333333

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant to the issue of notification availability, 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', which is unrelated to 'enabling notifications in the wiki settings'. The second node is about 'choosing notification email preferences and other customization options', which also does not address the issue of enabling notifications. The third node lists 'various notification preferences and macros', none of which relate to enabling notifications. The fourth node describes 'the notifications application and its features', without mentioning enabling notifications. Finally, the fifth node talks about 'getting an RSS feed for notifications and settings', again not mentioning enabling notifications in the wiki settings.

ContextualRecall: The score is 0.50 because while the retrieval context (node 5) supports the general idea of enabling notifications in the wiki, it lacks specific details about setting the 'notifications.enabled' property in 'xwiki.properties' to 'true', which is crucial for full alignment with the expected output.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.5171428571428571

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.79 because the response mostly addresses the issue of not receiving notifications, but includes irrelevant source links and references 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. The first node 'lists various features and settings related to notifications but does not provide specific instructions on enabling notifications or following users/pages', which is not helpful for the input. The second node 'discusses notification settings and options but lacks direct guidance on enabling notifications or following users/pages', which is crucial to the expected output. The third node 'explains the availability of RSS feeds and notification settings but does not address enabling notifications or following users/pages', which are key components of the expected output. Lastly, the fourth node 'provides a list of notification-related macros and preferences but does not offer specific instructions on enabling notifications or following users/pages', which are necessary for the expected output.

ContextualRecall: The score is 0.50 because while there are connections between the expected output and nodes in the retrieval context regarding enabling notifications and following users or pages, several specific details such as setting 'notifications.enabled' in 'xwiki.properties', the 'alert' menu, and viewing followed users in the 'network' tab are not supported by the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting 'notifications.enabled' in 'xwiki.properties' and specifics about following pages or viewing followed users in the user profile.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_023_result.json

Overall_score: 0.9514285714285714

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.86 because the response is mostly relevant to the question about enabling notifications for one's own actions, 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. 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 and directly addresses the input question about receiving notifications for one's own activity. Great job on ranking the relevant information at the top!

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 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, the specific location 'Options de filtrage avancées' is not explicitly mentioned.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.940952380952381

Individual_scores:

AnswerRelevancy: 0.9047619047619048

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.90 because the response effectively addresses the question about disabling notifications, but it includes irrelevant links that do not directly help with managing or 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 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant node. The first three nodes effectively address the issue of automatic notifications by discussing 'Choosing the default auto-watched mode', 'Notifications Auto Watch Preferences', and 'Auto Watch' under 'Watched Entities'. The irrelevant node, ranked fourth, discusses 'Get a notification RSS feed' and 'Notifications Macro', which do not directly relate to managing automatic notifications, ensuring a perfect score. Great job!

ContextualRecall: The score is 1.00 because all sentences in the expected output are well-supported by the nodes in the retrieval context, demonstrating a perfect alignment.

CustomContextualRelevancy: The retrieval context includes information about auto-watched pages and custom notification filters, aligning with the expected answer. However, it lacks explicit mention of modifying the 'Suivi automatique des pages' setting in user profile notification settings.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_025_result.json

Overall_score: 0.42666666666666664

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to configuring OpenID Connect in XWiki, but it includes a minor irrelevant detail about LDAP Authenticator, 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 'domain-based authentication and session cookies', which does not relate to enabling user accounts automatically with OpenID Connect. The second node is about 'container authentication' and only mentions OpenID Connect in passing, without addressing user account activation. The third node provides a 'general description of OpenID Connect authentication'

but lacks details on automatic user account activation. The fourth node lists 'dependencies for the OpenID Connect extension', unrelated to the needed configuration. The fifth node discusses 'LDAP and Active Directory configurations', which are not related to the OpenID Connect user account activation issue.

ContextualRecall: The score is 0.20 because only the 3rd node in the retrieval context aligns with the expected output by mentioning 'OpenId Connect Authentication' and the 'xwiki.properties' file, while other details about specific property updates, uncommenting, saving, and restarting are not supported by the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention adjusting the xwiki.properties file or the oidc.enableUser property for automatic user account activation via OpenID Connect.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.4514285714285714

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.86 because the answer is mostly relevant and addresses the question about customizing user profile fields in XWiki. However, it includes a repeated irrelevant statement about a source link that doesn't directly address the input question, which 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'authentication management class and OpenID Connect provider endpoints' without mentioning 'user profile field customization or mapping'. The second node is about 'container authentication and other authentication methods like LDAP and Active Directory', which is unrelated to the input query. The third node provides 'general information about the OpenID Connect Authenticator' but lacks details on 'user profile field customization or mapping'. The fourth node lists 'dependencies for the OpenID Connect Authenticator extension', which is not related to the input. Lastly, the fifth node mentions 'configuration guides and limitations' without addressing 'how to map user profile fields from OpenID Connect'.

ContextualRecall: The score is 0.40 because while the 1st node in the retrieval context supports the customization of user profile fields and saving changes in 'xwiki.properties' as mentioned in sentences 1 and 5 of the expected output, it lacks details on 'oidc.user.mapping', examples of mapping properties, and specific replacements mentioned in sentences 2, 3, and 4.

CustomContextualRelevancy: The retrieval context does not mention mapping user profile fields or using 'oidc.user.mapping' in 'xwiki.properties', which are key aspects of the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6066666666666667

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.4

CustomContextualRelevancy: 0.3

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, which discusses 'group synchronization, including the use of claims and mappings in the xwiki.properties file', is ranked third. This means it is ranked lower than two irrelevant nodes. The first node discusses 'authentication management and endpoint configurations', and the second node focuses on 'different authentication methods and extensions, such as LDAP and Active Directory', both of which are not directly related to group synchronization setup. These irrelevant nodes should be ranked lower to improve the score.

ContextualRecall: The score is 0.40 because while the 3rd node in the retrieval context supports configuring group synchronization, claims, separators, and mappings as mentioned in sentences 3, 5, and 7 of the expected output, it lacks details on locating, opening, and saving the 'xwiki.properties' file, as well as replacing placeholders with actual values 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 configuring oidc.groups.claim and oidc.groups.separator, which are specified in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.32857142857142857

Individual_scores:

AnswerRelevancy: 0.6428571428571429

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.64 because the output partially addresses the question about restricting file uploads to images or PDFs of a specific size, but includes multiple irrelevant source links that do not directly answer 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 'language preferences and translations in XWiki', which is unrelated to file upload restrictions. The second node describes 'various XWiki applications and features', but does not address file upload restrictions. The third node is about 'CSS and XSL for PDF styling in XWiki', which is not relevant to configuring file upload restrictions. The fourth node lists 'various XWiki macros and their functionalities', none of which pertain to setting file upload restrictions. The fifth node provides 'links to various XWiki references and images', which do not relate to configuring file upload settings. Lastly, the sixth node contains 'links to images and references', which are not relevant to the topic of file

upload restrictions.

ContextualRecall: The score is 0.00 because none of the expected output sentences are supported by the nodes in the retrieval context, as the context lacks information on file size restrictions, MIME type configuration, and specific configuration steps mentioned in the expected output.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.6900000000000001

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.75 because the response addresses the main question about refusing script rights to a space administrator, but it includes irrelevant source links that do not directly relate to the input 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 0.50 because the second node in the retrieval context is relevant, discussing the 'Script' right and its denial, which aligns with the input. However, the first node is irrelevant as it focuses on 'language preferences and translations in XWiki,' which should be ranked lower. Additionally, the third node, which 'lists various rights and their default states,' is also irrelevant and should be ranked lower than the relevant node.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the information in the retrieval context, showing a clear understanding of XWiki's rights management. 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 main point of the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.35789473684210527

Individual_scores:

AnswerRelevancy: 0.7894736842105263

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.79 because the answer mostly addresses the configuration question about XWiki permissions, but it includes multiple irrelevant source links that do not contribute to solving the issue, 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. For example, the first node is about 'the structure of an XWiki page and its metadata', which does not address access control or permissions. The second node discusses 'modifying resources and creating pages using XML and HTTP methods', unrelated to configuring access permissions in XWiki. The third node focuses on 'adding, editing, and deleting comments', which is not related to setting access permissions for spaces in XWiki. The fourth node explains 'the creation of spaces and sub-pages in XWiki', but does not provide information on setting access permissions for teams. The fifth node is about 'migration actions related to page renaming and code adjustments', which are not relevant to configuring access permissions in XWiki. Lastly, the sixth node discusses 'configuring display scopes for administration pages', which does not address the question of 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 any information about creating groups, setting permissions, or inheriting rights as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7495238095238095

Individual_scores:

AnswerRelevancy: 0.7333333333333333

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.73 because the answer addresses the question about authenticating users with access tokens, but includes multiple irrelevant source links that do not contribute to the explanation, 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 0.50 because the relevant nodes in the retrieval context are not consistently ranked higher than irrelevant nodes. The first node discusses 'LDAP authentication settings in XWiki', which is unrelated to token-based authentication, yet it is ranked higher than the second node, which is relevant as it provides information on configuring 'Authorized Applications'. Similarly, the third node, discussing 'OpenID Connect scopes and response types', is not directly related to token-based authentication, but is ranked higher than the fourth node, which directly addresses authentication using JWTs. However, the relevant nodes are still present, contributing to the score being 0.50.

ContextualRecall: The score is 0.71 because while several key aspects such as the use of JWTs, enabling the token authenticator, and configuring authorized applications are supported by nodes in the retrieval context, some details like generating a JWT token with specific claims and including the token in the 'Authorization' header are not covered.

CustomContextualRelevancy: The retrieval context contains information about configuring authorized applications, JWT token requirements, and enabling the token authenticator, aligning well with the expected answer. However, specific details about claims like expiration time and the use of Ed25519 key are not explicitly mentioned.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_032_result.json

Overall_score: 0.3555555555555555
Individual_scores:
AnswerRelevancy: 0.7777777777777778
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
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 pertain 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. The first node discusses 'model validation strategies in XWiki', the second node covers 'attachment validation and configuration in XWiki', and the third node provides details about 'the admin UI for model validation and its installation'. None of these are related to 'authenticating users with 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 mismatch.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWTs as described in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_033_result.json

Overall_score: 0.3
Individual_scores:
AnswerRelevancy: 0.5
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.50 because while the output attempts to address the question about authenticating users with access tokens, it includes irrelevant information such as unrelated source links and discussions about user registration settings and captchas, which do not directly pertain 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 token-based authentication'. The second node focuses on 'features of the Administration Application' related to 'users and rights', but lacks information on 'authentication with access tokens'. The third node mentions

'rights and authentication guides', yet fails to provide details 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 answer.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_mistral2_large

File: qa_001_result.json

Overall_score: 0.520995670995671

Individual_scores:

AnswerRelevancy: 0.8787878787878788

Faithfulness: 0.9166666666666666

ContextualPrecision: 0.30952380952380953

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the response mostly addresses the issue of the missing 'Bell' icon in XWiki, but the inclusion of multiple source links that do not directly address the problem slightly reduces its relevance.

Faithfulness: The score is 0.92 because the actual output suggests navigating to the Administration page and selecting the Import category to upload the XAR file for the Notifications Application, which is not specified in the retrieval context as the method for installing the Notifications Application.

ContextualPrecision: The score is 0.31 because relevant nodes in the retrieval context, such as the third node which 'mentions various notification methods in XWiki, including the Notifications Application,' and the seventh node which 'discusses how administrators can configure notification settings, including enabling notifications by default,' are ranked lower than irrelevant nodes. For example, the first node 'discusses installation instructions and prerequisites for extensions,' which is not relevant to enabling notifications, and should be ranked lower.

ContextualRecall: The score is 0.50 because while the 7th node in the retrieval context relates to enabling notifications, there is no mention of the specific setting 'notifications.enabled' in 'xwiki.properties'.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.705925925925926

Individual_scores:

AnswerRelevancy: 0.9629629629629629

Faithfulness: 0.8333333333333334

ContextualPrecision: 0.6333333333333333

ContextualRecall: 0.5

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.96 because the answer is highly relevant to the question about notifications in XWiki, with only a minor irrelevant mention of RSS features that doesn't significantly detract from the overall helpfulness.

Faithfulness: The score is 0.83 because the actual output incorrectly claims that new users in XWiki versions prior to 15.5 do not receive notifications until they take specific actions, contradicting the retrieval context which states that the whole wiki is watched for new users by default.

ContextualPrecision: The score is 0.63 because the first node in the retrieval context is relevant, explaining how to follow a user and receive notifications, ensuring notifications are set up correctly. However, the second node discusses filters and hidden pages, which do not directly relate to enabling notifications, and should be ranked lower. The third node is about customizing email templates for notifications, which is not directly related to enabling notifications, and should also be ranked lower. Additionally, the fourth node discusses RSS feeds, an alternative method of receiving notifications, but does not address the issue of not receiving notifications, indicating it should be ranked lower. The fifth node is relevant as it mentions the settings menu for notifications, directly related to ensuring notifications are enabled. The sixth node is also relevant, listing various ways to receive notifications, including the Notifications Application. However, the seventh node is about administrator settings for notifications, which is not directly relevant to a user not receiving notifications, and should be ranked lower. Lastly, the eighth node describes the Alerts menu and how notifications are displayed, which does not directly address the issue of not receiving notifications, indicating it should be ranked lower.

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 method to enable notifications via the 'alert' menu (node 6) and following a user for event notifications (node 1), other key details like enabling notifications in `xwiki.properties` and subscribing to pages or users are not supported by any nodes.

CustomContextualRelevancy: The retrieval context includes information on following users and watching pages, but lacks details on enabling notifications via 'xwiki.properties' and using the alert menu.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9400000000000001

Individual_scores:

AnswerRelevancy: 0.7

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.70 because the response provides relevant information about enabling notifications for one's own actions in XWiki, but it includes irrelevant source links that do not directly address the 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 1.00 because the relevant nodes in the retrieval context are ranked higher than the irrelevant ones. 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,' which directly aligns with the expected output. The second node mentions '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, which is about following other users, and the fourth node, which is about administrators configuring default notification settings, are ranked lower, ensuring a perfect score.

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 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 in the user profile.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6197680097680098

Individual_scores:

AnswerRelevancy: 0.9615384615384616

Faithfulness: 1.0

ContextualPrecision: 0.33730158730158727

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant and effectively addresses the question about disabling notifications in XWiki. The only minor issue is the inclusion of a source link, which doesn't directly answer the question but doesn't significantly detract 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 0.34 because the first three nodes in the retrieval context are irrelevant, discussing topics like 'following a user and receiving notifications about events triggered by that user', 'selecting the level of detail in email notifications', and 'enabling and disabling the email feature', which do not address how to disable notifications for pages. These irrelevant nodes are ranked higher than relevant nodes that discuss 'administrators changing users' notification settings', 'default settings for notifications', and 'filters for notifications', which are pertinent to the input query.

ContextualRecall: The score is 0.50 because only the information about default auto watching page behavior is supported by nodes in the retrieval context, specifically node 5. However, details about being notified of changes and deleting notification filters in the 'Custom Filters' list are not supported by the retrieval context.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.9066666666666666

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to configuring OpenID Connect in XWiki, but it includes a mention of LDAP Authenticator, which is unrelated to the question, 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 first node in the retrieval context is directly relevant, mentioning the '# oidc.enableUser=true' property, which is precisely about enabling user accounts on first login. All irrelevant nodes, such as those discussing 'domain-based authentication and session cookies' (second node), 'password reset and various authentication mechanisms' (third node), and others, are correctly ranked lower.

ContextualRecall: The score is 0.60 because while the 1st node in the retrieval context supports the need 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 confirming that new user accounts are automatically enabled upon first login via OpenID Connect.

CustomContextualRelevancy: The retrieval context includes the exact information about the 'oidc.enableUser' property being set to true and needing to be uncommented, which matches the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.925

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.625

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

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 0.62 because the first node in the retrieval context is relevant as it mentions 'oidc.user.mapping=myxproperty1=\${oidc.user.subject}', which directly relates to mapping additional user profile fields from the OpenID Connect provider to XWiki user properties. However, irrelevant nodes discussing 'access tokens and scopes' (second node), 'group synchronization and custom claims' (third node), 'authentication management class and OpenID Connect provider endpoints' (fourth node), and other unrelated topics are ranked higher than the second relevant node, which lists 'available variables such as oidc.user.subject' (eighth node). These irrelevant nodes should be ranked lower to improve the score.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the information in the 1st node in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context contains detailed information about customizing user profile fields using the 'oidc.user.mapping' property in the 'xwiki.properties' file, matching the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8189743589743591

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.8333333333333334

ContextualRecall: 0.46153846153846156

CustomContextualRelevancy: 0.8

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 discussing 'oidc.groups.mapping', 'oidc.groups.separator', and 'oidc.groups.claim' are ranked higher, providing pertinent details for configuring group synchronization. However, the third node primarily discusses authentication and general OpenID Connect setup, which is not specifically about group synchronization, and should be ranked lower. Similarly, the fourth node focuses on authentication management and endpoint configuration, and the fifth node lists dependencies, both of which are not directly related to the group synchronization setup needed.

ContextualRecall: The score is 0.46 because while some nodes in the retrieval context support configuring properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', key actions such as locating and editing the 'xwiki.properties' file, replacing placeholders, and restarting the XWiki instance are not covered by the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context includes information about configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', but lacks explicit steps for locating and editing 'xwiki.properties' and restarting XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6272727272727272

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.6363636363636364

CustomContextualRelevancy: 0.3

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 0.20 because the relevant node, which 'describes how to set limits on the maximum size of an attachment and mentions the configuration steps', is ranked fifth. This means that many irrelevant nodes, such as the first node which 'discusses linking to external files and

images, but does not address configuring upload restrictions or size limits', are ranked higher than the relevant node. These irrelevant nodes, including the second node that 'explains the structure of XWiki pages and available actions, but does not mention attachment upload restrictions or size limits', should be ranked lower to improve the score.

ContextualRecall: The score is 0.64 because while several sentences in the expected output align well with nodes in the retrieval context, particularly steps for configuring mimetypes and maximum attachment size (nodes 4 and 5), introductory and explanatory sentences lack direct matches in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions, but lacks specific instructions for configuring mimetypes in xwiki.properties and setting the size to 10MB as detailed in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.7652173913043478

Individual_scores:

AnswerRelevancy: 0.8260869565217391

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.83 because the answer is mostly relevant to the question about denying script rights to a space administrator, but it includes multiple 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 question. The first node discusses 'programming rights and the implications of admin rights' but does not address denying script rights specifically. The second node provides 'a list of rights and their default states' but does not mention the inability to deny script rights to a space administrator. Subsequent nodes similarly fail to address the specific issue of denying script rights to a space administrator, focusing instead on unrelated topics such as 'comment, edit, and delete rights' and 'rights conversion and excluded spaces.'

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 retrieval context confirms that rights implied by admin cannot be denied, aligning with the expected answer that admin rights cannot be denied in XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.6081231231231231

Individual_scores:

AnswerRelevancy: 0.8378378378378378

Faithfulness: 1.0

ContextualPrecision: 0.5694444444444444

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.84 because the output mostly addresses the configuration question in XWiki, but the repeated source links do not directly contribute to solving the problem, slightly reducing the relevancy.

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

ContextualPrecision: The score is 0.57 because relevant nodes in the retrieval context, such as the first node discussing 'rights inheritance and administration features' and the sixth node providing information on 'XWiki's permission system', are mixed with irrelevant nodes. For example, the second node, ranked higher than some relevant nodes, focuses on 'parent/child relationship and nested pages', which is not directly related to configuring access rights for teams. Similarly, the third node discusses 'string formatting and OpenID Connect properties', which is unrelated to the input question. These irrelevant nodes being ranked higher than some relevant nodes contribute to the score not being higher.

ContextualRecall: The score is 0.33 because only a few sentences in the expected output are supported by the nodes in the retrieval context. Specifically, sentences 2 and 4 are supported by nodes 6 and 1 in the retrieval context, respectively, while most other sentences, including 1, 3, and the concluding sentences, lack direct support from the retrieval context.

CustomContextualRelevancy: The retrieval context mentions setting rights and inheritance but lacks specific steps like creating groups for each team and setting space-level rights, which are crucial in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.425

Individual_scores:

AnswerRelevancy: 0.625

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.62 because the response partially addresses the issue of the bell not being available, but includes irrelevant links to documentation that do not directly 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For instance, the first node discusses 'custom notification filters and their behavior', but it does not address the issue of 'enabling notifications in the wiki settings'. Similarly, the second node lists 'various ways to receive notifications in XWiki', yet it also fails to explain how to enable notifications in the wiki settings. This pattern continues with all nodes, as they do not provide the needed information about enabling notifications in the wiki settings.

ContextualRecall: The score is 0.50 because only part of the expected output is supported by the nodes in the retrieval context. Specifically, the mention of configuring notifications aligns with the 6th node in the retrieval context, but the specific instruction about setting 'notifications.enabled' in 'xwiki.properties' is not found in the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3666666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the response is mostly relevant to the issue of not receiving notifications, but it includes unrelated source links that slightly detract from its 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node contains 'image links and references, which do not relate to enabling notifications or following users/pages in XWiki.' Similarly, the second node is 'a list of image links and references, unrelated to the topic of notifications or user/page following in XWiki.' The third node also 'contains image links and references, which are irrelevant to the issue of receiving notifications or following users/pages.' The fourth node is 'a list of image links and references, not providing any information on enabling notifications or following users/pages.' Lastly, the fifth node describes 'a security module and its ratings, which is unrelated to the topic of notifications or following users/pages 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 does not contain any relevant information about enabling notifications, following pages or users, or managing notifications as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.3625

Individual_scores:

AnswerRelevancy: 0.8125

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.81 because the response is mostly relevant to the question about enabling notifications for one's own actions. However, the inclusion of multiple source links that do not directly address the question 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 unrelated to notification settings. 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 and rights, which is not relevant to notification settings.' The third node covers 'importing office documents and other features, but does not mention notification settings.' The fourth node is about 'themes and page templates, which do not relate to notification settings.' Finally, the fifth node discusses 'color themes and access rights, which are not relevant to notification settings.' These irrelevant nodes are ranked higher than any relevant information, resulting in 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 mention notification settings or filters related to user activity, which are key aspects of the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.3333333333333333

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.67 because the answer partially addresses how to disable notifications, but it includes multiple irrelevant source links that do not directly relate to the question, which lowers the relevancy 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 example, the first node is about 'attaching files and does not relate to notification settings or disabling notifications', which should be ranked lower. Similarly, the second node discusses 'moving attachments, which is unrelated to notification settings', and should also be ranked lower. This pattern continues with all nodes, such as the third node providing 'a list of macros and their uses, which is not relevant to notification settings', indicating that the relevant nodes are not ranked higher than irrelevant ones.

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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.46571428571428575

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 0.8

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and mostly addresses the question about configuring XWiki for automatic user account activation with OpenID Connect. However, the mention of configuring endpoints and headers is slightly off-topic, preventing a perfect score.

Faithfulness: The score is 0.80 because the actual output incorrectly claims that the property `xwiki.authentication.createuser=empty` ensures automatic user account creation, while the retrieval context clarifies that this property must be set in the xwiki.cfg configuration file for user creation, not in the xwiki.properties file.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question. The first node discusses 'domain-based authentication and session cookie configuration', which does not relate to enabling user accounts upon first login with OpenID Connect. The second node mentions OpenID Connect but lacks information on automatically activating user accounts. The third node provides a general description of OpenID Connect authentication but does not address the specific configuration needed. The fourth node describes the OpenID Connect authenticator extension without mentioning automatic user activation. The fifth node lists OpenID Connect endpoints and configuration options but not the specific property needed. The sixth node lists dependencies for the OpenID Connect authenticator extension, which is not relevant. Lastly, the seventh node provides links to features and troubleshooting without mentioning the configuration for automatic user activation.

ContextualRecall: The score is 0.60 because while there is relevant information in the nodes in retrieval context about configuring OpenID Connect and properties in 'xwiki.properties', specific instructions such as adding or updating the property and restarting the XWiki instance are not mentioned.

CustomContextualRelevancy: The retrieval context does not contain information about adjusting the `xwiki.properties` file or the `oidc.enableUser` property, which are critical for the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.46923076923076923

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.1

Reasons:

AnswerRelevancy: The score is 0.85 because the response is mostly relevant to the input question about customizing user profile fields in XWiki. However, it includes source links that do not directly address the customization process, which slightly reduces 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. For example, the first node discusses 'container authentication and Active Directory', which are unrelated to 'customizing user profile fields from an OpenID Connect provider'. Similarly, the second node is about 'domain-based authentication and session cookies', which does not relate to 'mapping user profile fields in XWiki'. None of the nodes address the specific need to customize or map user profile fields from an OpenID Connect provider in XWiki.

ContextualRecall: The score is 0.40 because only some parts of the expected output, specifically sentence 1 and part of sentence 2, can be linked to the 7th node in the retrieval context, which mentions 'xwiki.properties' and 'oidc.user.mapping'. However, the specific examples and instructions in sentences 3 and 4 are not supported by the retrieval context.

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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.44000000000000006

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

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. Great job on maintaining complete 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 setting up group synchronization between XWiki and an OpenID Connect provider. The first node discusses general OpenID Connect provider configuration and endpoints without mentioning group synchronization. The second node provides information on different authentication methods and extensions but does not address group synchronization. The third node focuses on domain-based authentication and session cookie configuration, which is unrelated to the group synchronization setup. The fourth node mentions automatic synchronization but lacks specific details on configuring group synchronization properties like 'oidc.groups.claim' or mapping. The fifth node lists dependencies for the OIDC authenticator extension, which is not relevant to the group synchronization configuration process. The sixth node includes a link to group synchronization but does not provide the detailed configuration steps needed. Lastly, the seventh node provides a brief mention of group synchronization but lacks detailed instructions or properties configuration necessary for the expected output.

ContextualRecall: The score is 0.20 because only a small portion of the expected output is supported by the nodes in the retrieval context. Specifically, sentence 3 is partially supported by node 1, and sentence 5 is partially supported by node 7. However, the majority of the expected output, including key steps and configurations, lacks direct support from the retrieval context nodes.

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', as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6433333333333333

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0.26666666666666666

ContextualRecall: 0.9

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.75 because the main content of the output addresses the input question about configuring upload restrictions effectively, 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 0.27 because the relevant nodes in retrieval context, specifically the fifth and sixth nodes, which mention 'Size Limit' and 'Mimetype Restriction', are ranked lower than several irrelevant nodes. The first four nodes discuss topics such as 'trusted domains and PDF export customization', 'ODT and PDF export features', 'WebDAV and PDF export', and 'PDF and RTF export issues', none of which relate to configuring upload restrictions for images and PDFs. These irrelevant nodes should be ranked lower than the relevant nodes to improve the score.

ContextualRecall: The score is 0.90 because most sentences in the expected output have corresponding nodes in the retrieval context, such as configuring MIME types and maximum attachment size, which are supported by nodes 1, 5, and 6. However, the introductory statement about uploading files lacks direct support in the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.36666666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 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 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. For instance, the first node discusses 'form validation and scripting', which is unrelated to denying script rights in XWiki. The second node is about 'template providers and page creation', not about rights management in XWiki. Similarly, the third node covers 'custom configuration in XWiki', which does not relate to denying script rights. The fourth node is about 'field validation in registration forms', unrelated to script rights management. Finally, the fifth node discusses 'access rights and scripting API', but does not address the specific issue of denying script rights to a space administrator in XWiki.

ContextualRecall: The score is 0.00 because the expected output discusses rights management in XWiki, but none of the nodes in the retrieval context mention rights or admin rights specifically, leading to a complete mismatch.

CustomContextualRelevancy: The retrieval context does not contain any information about XWiki not supporting the denial of rights implied by admin rights, which is the key fact in the expected answer.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.35

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.75 because the response partially addresses the question about configuring permissions in XWiki, but includes several irrelevant source links that do not directly relate to the specific task of setting permissions for team areas.

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 compatibility', which does not relate to configuring access controls for teams. The second node is about 'notifications and application details', not about configuring team access or editing rights in XWiki. The third node describes 'an administration application for managing wiki features', but lacks guidance on configuring team access rights. Subsequent nodes continue to be unrelated, discussing topics like 'Admin UI for Model Validation', 'Nested Pages and subwikis', 'managing users and groups', 'LDAP group configurations', 'group synchronization and customization options', and 'LDAP group mapping and Java code for authentication', none of which address the specific requirement of setting view and edit rights for different teams in XWiki.

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, indicating a complete lack of relevant information.

CustomContextualRelevancy: The retrieval context does not contain any relevant information about creating groups, setting rights, or access control in XWiki as outlined in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.8375

Individual_scores:

AnswerRelevancy: 0.9375

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant, addressing the main issue about the notification bell. However, it includes a minor irrelevant detail about XWiki syntax, which slightly affects 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 0.25 because the relevant node in the retrieval context, which includes a section on enabling notifications by setting 'notifications.enabled = true' in xwiki.properties, is ranked fourth. This node directly relates to the input, but it is ranked lower than irrelevant nodes. The first node discusses custom notification filters and their behavior, which does not address enabling notifications in the wiki settings. The second node lists various notification preferences and macros, which also does not mention enabling notifications. The third node talks about filtering notifications and hidden pages, which is unrelated to enabling notifications. 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 fully supported by the nodes in the retrieval context, specifically node 4, which provides all the necessary details about enabling notifications in the wiki.

CustomContextualRelevancy: The retrieval context explicitly mentions enabling notifications by setting 'notifications.enabled' in 'xwiki.properties' to 'true', which directly matches the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.7072222222222222

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0.7222222222222222

ContextualRecall: 0.625

CustomContextualRelevancy: 0.3

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 0.72 because relevant nodes are generally ranked higher than irrelevant ones. However, the score is not higher because the second node, which focuses on 'customizing email templates and creating global notification filters', and the fourth node, which discusses 'RSS feeds and notification grouping strategies', are ranked above the sixth node, which is relevant as it mentions 'Administrators can go to Administration > Notifications to configure settings'.

ContextualRecall: The score is 0.62 because while several elements of the expected output, such as enabling notifications and following pages or users, are supported by nodes in the retrieval context, specific details like 'xwiki.properties', 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' in 'xwiki.properties' and specific steps for following pages or users.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8355555555555556

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.78 because the response partially addresses the question about enabling notifications for one's own actions, but includes irrelevant 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 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 correctly ranked first. This directly addresses the input question about enabling notifications for one's own actions. Great job on getting it right!

ContextualRecall: The score is 0.50 because the 1st sentence in the expected output is supported by the 1st node in the retrieval context, but the 2nd sentence lacks support as the retrieval context does not mention 'Options de filtrage avancées' or related settings.

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 for disabling the filter.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.5678947368421052

Individual_scores:

AnswerRelevancy: 0.7894736842105263

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.79 because the response partially addresses the question about disabling notifications, but includes several irrelevant links and references that do not directly help in solving 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 'technical aspects of notification filters and settings' but fails to address disabling notifications or auto-watching pages. The second node lists 'various notification preferences and macros' without providing relevant information on disabling notifications. The third node offers 'links to various sections about notifications' but does not explain how to disable them. The fourth node talks about 'RSS feeds and notification settings' without addressing the disabling of notifications. Lastly, the fifth node outlines 'different notification methods like RSS feeds' but lacks guidance on disabling

notifications or auto-watching pages.

ContextualRecall: The score is 0.75 because sentences 1 and 3 are supported by nodes in the retrieval context, specifically nodes 4 and 3, which relate to automatic notifications and modifying settings. However, sentence 2 lacks support as the retrieval context does not mention notifications about changes to one's work.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.4383333333333335

Individual_scores:

AnswerRelevancy: 0.9583333333333334

Faithfulness: 0.8333333333333334

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant to configuring OpenID Connect in XWiki, with only a minor mention of LDAP authentication, which slightly detracts from the focus but does not significantly impact the overall relevance.

Faithfulness: The score is 0.83 because the actual output incorrectly states that the authentication class for OpenID Connect must be defined in the xwiki.properties file, which is 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 'domain-based authentication and session cookie configuration', which does not pertain to enabling user accounts automatically via OpenID Connect. The second node focuses on 'OpenID Connect provider URLs and endpoints', not on the automatic activation of user accounts. Subsequent nodes similarly fail to address the specific configuration needed for automatic user account activation, such as discussing 'general information about OpenID Connect authentication', 'dependencies for the OpenID Connect extension', or 'LDAP and Active Directory integration', none of which are relevant to the input query.

ContextualRecall: The score is 0.40 because while the 2nd node in the retrieval context supports the mention of 'xwiki.properties' and 'oidc.enableUser' in relation to OpenID Connect configuration (supporting sentences 1 and 2), it lacks specific instructions on adding or updating properties, saving the file, restarting the instance, and the automatic activation of new user accounts upon first login (unsupportive for sentences 3, 4, and 5).

CustomContextualRelevancy: The retrieval context does not contain any information about setting the 'oidc.enableUser' property in 'xwiki.properties' to automatically activate user accounts on first login.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.550909090909091

Individual_scores:

AnswerRelevancy: 0.9545454545454546

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant to the question about mapping user profile fields in XWiki using OpenID Connect, with only a minor mention of UI validation 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. For example, the first node discusses 'authentication methods like Container Authentication and Active Directory,' which does not relate to 'mapping OpenID Connect user profile fields to XWiki properties.' Similarly, the second node focuses on 'domain-based instance authentication and session cookie configurations,' which are also unrelated to the input query.

ContextualRecall: The score is 0.60 because while there are relevant mentions in nodes in the retrieval context about customizing user profile fields and using the xwiki.properties file, there is a lack of specific examples and instructions on mapping properties like 'myxproperty1' or replacing them with actual XWiki user property names.

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.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5164473684210527

Individual_scores:

AnswerRelevancy: 0.39473684210526316

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.4375

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.39 because the output contains numerous code import statements and implementation details that do not directly address the process of configuring group synchronization between XWiki and an OpenID Connect provider. These irrelevant elements significantly lower the relevancy 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 0.25 because the relevant node in the retrieval context, which provides information on group synchronization, is ranked fourth. The first three nodes are irrelevant as they discuss 'authentication management and endpoints', 'container authentication', and 'domain-based authentication and session cookies', which do not address group synchronization or OpenID Connect group claims. This lowers the score as the relevant information is not prioritized.

ContextualRecall: The score is 0.44 because while the 4th node in 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 and editing the 'xwiki.properties' file, providing examples, and saving changes as required in sentences 1, 2, 4, 6, and 8.

CustomContextualRelevancy: The retrieval context mentions group synchronization and oidc.groups.mapping, but lacks details on oidc.groups.claim and oidc.groups.separator configuration steps.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.352

Individual_scores:

AnswerRelevancy: 0.76

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.76 because the main content of the output addresses the question about configuring file upload restrictions effectively, but the repeated irrelevant source links lower 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 'language preferences and translations in XWiki', which is unrelated to configuring file upload restrictions. The second node talks about 'various XWiki applications and features, such as tour guides and administration', which also does not mention file upload settings. The third node is about 'customizing PDF exports using CSS and XSL', which is not relevant to the question of file upload restrictions. All subsequent nodes continue to provide irrelevant information, such as 'trusted domains and PDF export customization', 'ODT export and PDF export features', and 'various XWiki features and macros, such as notifications and PDF viewers'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output are supported by the nodes in the retrieval context. The retrieval context lacks any mention of file size restrictions, MIME type configurations, or specific instructions related to editing XWikiPreferences, which are all crucial elements of the expected output.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.6854545454545455

Individual_scores:

AnswerRelevancy: 0.7272727272727273

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.73 because the main content addresses how to refuse script rights to a space administrator, but multiple irrelevant source links are included, which 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 nodes. For example, the first node is about 'language preferences and translations in XWiki,' which is unrelated to the question about script rights. Similarly, the third node discusses 'interacting with the XWiki RESTful API,' which is also irrelevant. However, the second node correctly addresses 'programming and script rights,' making it relevant to the input question, which is why the score is not lower.

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 mentions that admin rights imply other rights and are not deniable, but it does not explicitly state that admin rights cannot be refused as in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.36153846153846153

Individual_scores:

AnswerRelevancy: 0.8076923076923077

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.81 because the main content effectively addresses the question about configuring XWiki, but it includes multiple source links that do not directly contribute to answering 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'the metadata of a page in XWiki and does not provide information on configuring access rights for teams', which does not address the input query. Similarly, the second node talks about 'modifying resources using XML and HTTP methods', which is unrelated to setting access permissions in XWiki. All subsequent nodes also fail to provide relevant information on configuring 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.

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, which are all key components of the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.8333333333333334

Individual_scores:

AnswerRelevancy: 0.7666666666666667

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.77 because the main content addresses the question about authenticating users with access tokens, but the inclusion of multiple source links that do not directly contribute to the answer reduces the overall relevancy.

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 relevant nodes in the retrieval context are not consistently ranked higher than irrelevant nodes. For example, the first node discusses 'LDAP configuration and authentication', which is unrelated to token-based authentication, yet it is ranked higher than the second node that provides information on configuring 'Authorized Applications', which is relevant to using JWTs for authentication. Similarly, the third node is about 'LDAP authentication', which is not relevant, but it appears before the fourth node that details 'requirements for JWT token generation', directly relevant to the input. These misplacements contribute to the score not being higher.

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 detailed information about configuring authorized applications, JWT token generation, and claims required, matching the expected answer. However, it lacks explicit mention of the limitation regarding existing users not created through the authenticator.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.3

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

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 limits its relevance. However, it still partially addresses the question, hence 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 'Model Validation' without mentioning 'token-based authentication or JWT', which should be ranked lower. The second node talks about 'Attachment Validation and configuration of mimetype restrictions', unrelated to 'user authentication with access tokens'. The third node provides information about the 'Admin UI for Model Validation', not related to 'authentication or access tokens'. The fourth node is about 'setting up fields for user registration', not relevant to 'authentication using access tokens'. Lastly, the fifth node describes 'steps for installing a UI' and 'credentials for an Administrator', but does not relate to 'token-based authentication'.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be linked to any information in the 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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.275

Individual_scores:

AnswerRelevancy: 0.625

Faithfulness: 0.75

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.62 because the response partially addresses the question about authenticating users with access tokens, but the inclusion of irrelevant source links that do not directly pertain to the topic lowers the score.

Faithfulness: The score is 0.75 because the actual output incorrectly states that XWiki uses authentication APIs to manage user access rights, 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 is about 'upgrading XWiki' and does not mention 'user authentication with access tokens'. The second node focuses on 'features of the Administration Application' and does not address 'authentication with access tokens'. The third node mentions 'user authentication' but lacks information about 'using access tokens for authentication'. The fourth node discusses 'managing user rights' without mentioning 'authentication methods, such as using access tokens'. The fifth node lists 'various rights constants and their properties', which is unrelated to 'the authentication process 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. The retrieval context lacks information on authenticating users with access tokens, JWTs, or any related configurations and limitations mentioned in the expected output.

CustomContextualRelevancy: The retrieval context lacks any information about token-based authentication or JWT, which are central to the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_mixtral-8x22b

File: qa_001_result.json

Overall_score: 0.3576086956521739

Individual_scores:

AnswerRelevancy: 0.9130434782608695

Faithfulness: 0.875

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.91 because the response is largely relevant and helpful, addressing the issue of the bell icon not appearing in XWiki. However, it includes repeated irrelevant source links that do not directly solve the problem, preventing a perfect score.

Faithfulness: The score is 0.88 because the actual output inaccurately describes a manual method for accessing the Alert Notifications application, whereas the retrieval context specifies the manual method pertains to installing extensions instead.

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 help in enabling notifications. The second node describes 'the notifications menu and its features' without addressing how to enable notifications. The third node lists 'dependencies related to the notifications extension' but fails to resolve the enabling issue. Lastly, the fourth node talks about 'overriding notification templates and installation instructions', which are unrelated to the problem of enabling 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, as there is no mention of enabling notifications or the 'notifications.enabled' setting 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.7468253968253968

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8571428571428571

ContextualPrecision: 0.8055555555555555

ContextualRecall: 0.5714285714285714

CustomContextualRelevancy: 0.5

Reasons:

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

Faithfulness: The score is 0.86 because the actual output claims that new users won't receive notifications until they start watching something and enable some application types, which contradicts the retrieval context indicating that in versions before 15.5, the whole wiki is watched for new users by default.

ContextualPrecision: The score is 0.81 because the relevant nodes in the retrieval context are mostly ranked higher than the irrelevant ones. The first node is relevant as it explains how to follow a user to receive notifications, aligning with the expected output. However, the second node, which discusses filters and settings related to hidden pages and user activities without directly addressing notifications, is ranked higher than the third node. The third node is relevant as it discusses notification settings and enabling specific types of notifications, which should be ranked higher. Additionally, the fifth node, which talks about the 'What's New' application for news notifications, is not directly relevant to the expected output and should be ranked lower than the relevant fourth node that provides an overview of various notification methods in XWiki.

ContextualRecall: The score is 0.57 because while several aspects of the expected output, such as enabling notifications via the alert menu, following a user, and using the network tab, are supported by the 1st node in the retrieval context, key details like enabling notifications in 'xwiki.properties',

subscribing to pages or users, and watching a page or wiki are not covered in the nodes.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 1.0

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response was perfectly relevant and directly addressed 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. 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 ones. 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,' directly addressing the query. 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. Irrelevant nodes, such as the third node, which discusses administrative settings and notification preferences at a broader level, 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 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' of the notification settings.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.85

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.95

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant with no irrelevant statements, directly addressing the question about disabling 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 0.95 because the relevant nodes in the retrieval context are mostly ranked higher than the irrelevant ones. However, the fourth node, which 'primarily discusses email notifications and server settings,' and the sixth node, which 'is about overriding notification templates and installation instructions,' are not directly relevant to the user's query about disabling notifications for unwanted pages and should be ranked lower than the relevant nodes.

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

CustomContextualRelevancy: The retrieval context mentions default notification settings and user profile settings but lacks specific details about automatic page watching and deleting notification filters as described in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8457142857142858

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to the question about configuring OpenID Connect in XWiki, but it includes an unrelated 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 commented property '# oidc.enableUser=true' as directly relevant to enabling user accounts on first login, is ranked first. This ensures that the most pertinent information is prioritized, leading to a perfect score. Great job!

ContextualRecall: The score is 0.40 because while the 1st node in the retrieval context supports the need to define and set the 'oidc.enableUser' property, it lacks specific instructions on adding or updating the property, saving the file, restarting the XWiki instance, and the automatic enabling of new user accounts.

CustomContextualRelevancy: The retrieval context contains the necessary 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.9433333333333334

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.8

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant and 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 question, 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 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 about customizing user profile fields. The irrelevant nodes, discussing topics like 'scopes requested from the OpenID Connect provider', 'configuration of OpenID Connect endpoints and custom HTTP headers', 'group synchronization and customization of templates and listeners', and 'in-wiki configuration for the OIDC authenticator', are all ranked lower, ensuring perfect precision. Great job on the ranking!

ContextualRecall: The score is 0.80 because most of the expected output is supported by the 1st node in the retrieval context, including the customization of user profile fields and the example mappings. However, the instruction to save and restart XWiki is not explicitly mentioned in the retrieval context, leading to a slightly lower score.

CustomContextualRelevancy: The retrieval context contains all the information found in the expected answer, including the customization of user profile fields using the `oidc.user.mapping` property in the `xwiki.properties` file and examples of mapping custom properties.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8300000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.75

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

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

Faithfulness: The score is 0.75 because the actual output incorrectly claims that the `oidc.groups.claim` property is used to specify the claim for retrieving the list of groups from the OpenID Connect provider, while the retrieval context clarifies that this property is used to indicate a non-standard field for the list of groups.

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 information on configuring the `oidc.groups.claim`, `oidc.groups.mapping`, and `oidc.groups.separator` properties, which are essential for group synchronization. The irrelevant nodes, ranked fourth and fifth, discuss topics like authentication and configuration guides, and authentication management and endpoints, which are not directly related to the group synchronization process.

ContextualRecall: The score is 0.50 because while nodes in the retrieval context provide specific examples and mention key properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, they lack general instructions found in the expected output, such as locating and

editing the 'xwiki.properties' file or restarting the XWiki instance.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6290821256038648

Individual_scores:

AnswerRelevancy: 0.9565217391304348

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.5555555555555556

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant and effectively addresses the question about limiting attachment types and sizes. The minor irrelevant statement about XWiki basic concepts slightly detracts from a perfect score, but overall, the answer remains focused and useful.

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 retrieval context, which provides information on setting maximum attachment size and mentions the XWikiPreferences document, is ranked third. However, irrelevant nodes are ranked higher, such as the first node discussing linking to external files and image display, and the second node explaining the basic structure of XWiki pages, both of which do not address attachment upload settings or restrictions.

ContextualRecall: The score is 0.56 because while the retrieval context's 3rd node provides detailed steps for configuring the maximum attachment size (supporting sentences related to Step 2), it lacks information on configuring allowed mimetypes in the xwiki.properties file, which is crucial for Step 1.

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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.9904761904761905

Individual_scores:

AnswerRelevancy: 0.9523809523809523

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant to the question about denying script rights to a space administrator, with only a minor irrelevant mention of content

organization and 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 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first three nodes in the retrieval context provide pertinent information about the inability to deny script rights to a space administrator, such as 'programming rights can only be granted from the wiki preferences page' and 'ADMIN rights are not deniable.' The irrelevant nodes, ranked fourth and fifth, discuss unrelated topics like 'checking access rights programmatically' and 'converting rights during page migration,' which do not address the input query.

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

CustomContextualRelevancy: The expected output is fully supported by the retrieval context, which states that admin rights imply other rights and cannot be denied, matching the expected answer.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.755

Individual_scores:

AnswerRelevancy: 0.7333333333333333

Faithfulness: 1.0

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.875

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.73 because the response provides relevant information on configuring permissions in XWiki, but the inclusion of multiple source links that do not directly address the configuration 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 0.67 because the relevant nodes in the retrieval context are generally ranked higher than the irrelevant ones. However, the second node focuses on the 'parent/child relationship and the introduction of nested pages', which does not directly address configuring view and edit permissions for teams, and it is ranked higher than a relevant node that explains the XWiki permission system. Additionally, the third node is about 'string formatting and OpenID Connect properties', which is unrelated to configuring access rights in XWiki, yet it is ranked higher than another relevant node that provides basic rules about setting wiki-wide and page-level rights.

ContextualRecall: The score is 0.88 because most of the expected output sentences are supported by the nodes in the retrieval context, particularly node 4, which covers aspects like creating groups, setting wiki-wide and space-level rights, and rights inheritance. However, the retrieval context lacks specific information on 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, aligning partially with the expected answer. However, it lacks specific details on creating groups for each team and setting 'Edit' rights for team spaces.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.49333333333333335

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.80 because the response largely addresses the issue of the unavailable bell icon, but includes 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 0.17 because the relevant node, ranked sixth, discusses 'editing the xwiki.properties file, which is directly related to enabling notifications by setting `notifications.enabled` to `true`.' However, it is ranked lower than irrelevant nodes. For example, the first node discusses 'custom notification filters and their behavior, which is not directly related to enabling notifications in the wiki settings,' and should be ranked lower.

ContextualRecall: The score is 0.50 because while there is a connection between configuring settings in the retrieval context (node 5) and the expected output sentence about editing `xwiki.properties`, there is no specific mention of setting `notifications.enabled` to `true`, leading to incomplete alignment.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.36923076923076925

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses the issue of not receiving notifications, but it includes irrelevant details about sources and XWiki skins, which slightly detract 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 contains 'a list of image links and references, which are unrelated to the topic of enabling notifications or following users/pages for notifications.' The second node also 'contains image links and references, which do not provide any information relevant to configuring notifications or following entities for updates.' Similarly, the third node 'includes image links and references, offering no guidance on notification settings or following users/pages.'

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 only contain image links and references, not the detailed instructions or settings mentioned.

CustomContextualRelevancy: The retrieval context contains image links and references unrelated to notifications or user settings, lacking any relevant information from the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_013_result.json

Overall_score: 0.32
Individual_scores:
AnswerRelevancy: 0.6
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.60 because the response contains several irrelevant links that do not directly address the question about enabling notifications for one's own activity. However, it partially addresses the topic, which is why it 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 query. The first node discusses 'customizing page templates, internationalization, and other XWiki features', the second node covers 'the administration application for managing wiki features', and the third node talks about 'importing office documents, syntax highlighting, and other XWiki extensions'. None of these are related to enabling notifications for one's own activity, 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 about notification settings or user activity filters as mentioned in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_014_result.json

Overall_score: 0.3375
Individual_scores:
AnswerRelevancy: 0.6875
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.69 because the output partially addresses the question about disabling notifications, but includes multiple 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'attaching files and related features', which does not relate to 'managing notification settings'. Similarly, the second node is about 'moving attachments', which is unrelated to 'disabling notifications'. All nodes fail to address the issue of notification settings, 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 lacks any information about automatic page following, notification settings, or custom filters mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.4166666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.25

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the question about configuring XWiki for automatic user activation with OpenID Connect, but includes a workaround suggestion that doesn't directly solve the issue, slightly lowering 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 question. The first node discusses 'domain-based authentication and session cookie configuration,' which does not relate to 'automatic user activation in XWiki using OpenID Connect.' The second node provides 'general information about OpenID Connect authentication and configuration guides,' but lacks specifics on 'automatically activating user accounts.' The third node describes 'OpenID Connect authenticator and its features' without mentioning 'the specific configuration change needed in the xwiki.properties file.' Finally, the fourth node lists 'various OpenID Connect endpoints and configuration options' but omits 'the `oidc.enableUser` property or how to activate user accounts automatically.'

ContextualRecall: The score is 0.25 because only the instruction to update properties in 'xwiki.properties' is supported by node 4 in the retrieval context, while key details about 'oidc.enableUser' and automatic activation of user accounts are not covered in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain information about automatically activating user accounts with OpenID Connect in XWiki by modifying the xwiki.properties file.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.375

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the answer mostly addresses the question about customizing user profile fields in XWiki, but it includes irrelevant source links that don't directly help in achieving the task.

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 'authentication management and OpenID Connect provider configuration' without mentioning 'user profile field mapping or customization'. The second node is about 'container authentication and other authentication methods', lacking 'specific information on mapping OpenID Connect user profile fields in XWiki'. The third node provides a 'general description of the OpenID Connect Authenticator and its installation', not addressing 'user profile field customization or mapping'. The fourth node includes 'metadata about the OpenID Connect Authenticator', without mentioning 'user profile field mapping'. The fifth node lists 'dependencies for the OpenID Connect Authenticator extension', which is unrelated to 'customizing user profile fields from an OpenID Connect provider'.

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 customizing user profile fields using 'oidc.user.mapping' in 'xwiki.properties' as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.37538461538461537

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.07692307692307693

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.80 because the answer largely addresses the setup process for group synchronization between XWiki and an OpenID Connect provider. However, it loses some points due to a statement that mentions a lack of information about the group synchronization setup, which doesn't directly answer 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 'general OpenID Connect provider configuration and endpoints' but lacks information on group synchronization. The second node mentions 'authentication methods and extensions' without addressing group synchronization setup. The third node provides 'links to configuration guides and examples' but omits specific instructions for group synchronization. The fourth node lists 'dependencies for the OpenID Connect authenticator extension' without covering group synchronization setup. Finally, the fifth node describes 'the OpenID Connect authenticator's features and compatibility' but does not include details on group synchronization.

ContextualRecall: The score is 0.08 because only the mention of the 'xwiki.properties' file in the 1st node in retrieval context supports the expected output, while all other details related to setting up group synchronization, configuration properties, and examples are missing in 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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.5735714285714285

Individual_scores:

AnswerRelevancy: 0.6428571428571429

Faithfulness: 1.0

ContextualPrecision: 0.325

ContextualRecall: 0.6

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.64 because the output contains some relevant information about upload restrictions, but several irrelevant source links dilute its focus on the input 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 0.33 because the first three nodes in the retrieval context are irrelevant, discussing topics such as 'trusted domains and PDF export customization', 'ODT and PDF export features', and 'various macros and functionalities', which do not pertain to configuring upload restrictions for images or PDFs. The relevant nodes, which mention 'Size Limit', 'Mimetype Restriction', and provide steps to set the maximum size of an attachment, are ranked fourth and fifth. Additionally, the sixth node is irrelevant, focusing on 'document lifecycle features'. This misordering of relevant and irrelevant nodes results in a lower score.

ContextualRecall: The score is 0.60 because while the 5th node in retrieval context supports several steps related to configuring maximum upload size in XWikiPreferences (sentences 2-6), there is a lack of information regarding MIME type configuration and the restriction to images and PDFs (sentence 1).

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.3675925925925926

Individual_scores:

AnswerRelevancy: 0.9629629629629629

Faithfulness: 0.875

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant and directly addresses the question about denying script rights to a space administrator. However, it includes source links that do not contribute to answering the question, slightly reducing the score.

Faithfulness: The score is 0.88 because the actual output incorrectly specifies creating an XSheet named 'Custom Right Check', whereas the retrieval context only refers to a sheet contributing dt and dd elements without mentioning an XSheet.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'validation and programming aspects', which are unrelated to the question about denying script rights in XWiki. The second node is about 'template recommendations and terminal page creation', which does not relate to script rights or their denial in XWiki. The third node involves 'custom configuration and code related to XWiki classes', but it does not address the issue of 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 node in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information about XWiki not supporting the denial of rights implied by admin rights, as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.38655462184873945

Individual_scores:

AnswerRelevancy: 0.6470588235294118

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2857142857142857

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.65 because while the output attempts to address the question about configuring team access and editing permissions in XWiki, it includes several irrelevant links and statements that do not directly relate to the specific configuration task, thus lowering 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 'databases and programming features in XWiki,' which does not relate to configuring team access and editing permissions. The second node is about 'notifications and application details,' which are not pertinent to the task of setting access or editing permissions for teams. The third node explains 'using nested pages and subwikis,' which is not directly related to the configuration of access and editing rights for teams. The fourth node provides a link to 'managing users and groups,' but it does not directly address setting specific access or editing permissions for teams. Lastly, the fifth node discusses 'group synchronization and customization,' which does not directly relate to configuring team access and editing permissions in XWiki.

ContextualRecall: The score is 0.29 because only sentences 1 and 4 from the expected output are supported by nodes in the retrieval context, specifically nodes 4 and 3. The majority of the sentences, including those about achieving access control, setting wiki-wide and area-level rights, summarizing the steps, and adjusting rights for additional teams, lack direct support from the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.37333333333333335

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.87 because the response mostly addresses the issue of the notification bell not being available, but includes irrelevant links that do not directly resolve 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 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 customization options' but does not mention enabling notifications in the wiki properties. The third node lists 'different notification preferences' but does not mention enabling notifications through the 'xwiki.properties' file. The fourth node describes 'the alert menu and notification display' but does not provide information on enabling notifications in the wiki settings. The fifth node discusses 'RSS feeds and notification settings' but does not mention enabling notifications in the 'xwiki.properties' file.

ContextualRecall: The score is 0.00 because the expected output sentences about activating notifications and setting 'notifications.enabled' in 'xwiki.properties' are not supported by any nodes in the retrieval context, which focus on custom filters and notification preferences without mentioning these specific steps.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.775

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.625

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.75 because the response partially 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 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 crucial for ensuring notifications are activated. The second node includes 'Follow a user' and 'Watched Entities,' directly relating to receiving notifications. Irrelevant nodes, such as the third node focusing on 'RSS feeds and notification settings' and the fourth node listing 'various notification preferences macros,' do not address enabling notifications or following users/pages and are correctly ranked lower.

ContextualRecall: The score is 0.62 because while several nodes in the retrieval context support the expected output sentences about enabling notifications, following pages or users, and receiving notifications from followed users (as seen in sentences 1, 3, 4, and 5), there are missing details in the retrieval context regarding specific configurations and user interface elements mentioned in sentences 2, 6, and 7.

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.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8577777777777778

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant to the question about enabling notifications for one's own actions, but it includes an irrelevant link to XWiki syntax documentation, which 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 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 topics like RSS feeds, general notification settings, and unrelated links, are ranked lower.

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 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 option to change this setting. However, it does not explicitly mention the 'Options de filtrage avancées' section.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.9028571428571428

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about disabling notifications, but includes irrelevant details such as unrelated source links and discussions about notification timing, which do not directly help in 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 node. The irrelevant node, ranked fourth, discusses 'RSS feeds and settings for notification types', which are not directly related to disabling notifications or managing auto-watched pages. Great job on maintaining perfect precision!

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 includes information on auto-watched pages and custom notification filters, aligning with the expected answer, but lacks explicit mention of modifying settings for 'Suivi automatique des pages'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.40499999999999997

Individual_scores:

AnswerRelevancy: 0.625

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.62 because the output attempts to provide sources that might help with configuring XWiki for automatic user activation, but the URLs themselves do not directly address the specific question, leading to a moderate level of 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 'domain-based authentication and session cookies,' which is unrelated to enabling user accounts automatically with OpenID Connect. The second node mentions OpenID Connect but focuses on 'container authentication' without providing information on automatic user activation. The third node talks about 'installation, dependencies, and limitations,' not on configuring automatic user activation. The fourth node lists 'dependencies for the OpenID Connect extension,' which is not relevant to the input query. The fifth node is about 'LDAP and Active Directory integration,' which does not pertain to OpenID Connect or automatic user activation in XWiki.

ContextualRecall: The score is 0.40 because while the nodes in the retrieval context mention the 'xwiki.properties' file and modifying properties related to OpenID Connect, they do not specifically address adding or updating the property for user account activation, saving the file, restarting XWiki, or guaranteeing automatic activation of new user accounts.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.4492307692307692

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.85 because the response mostly addresses the question about customizing profile fields in XWiki with OpenID Connect. However, it includes irrelevant information about sources, which does not directly help in 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 'authentication management class and OpenID Connect provider endpoints' without mentioning 'user profile field mapping or customization'. The second node talks about 'container authentication and Active Directory', which does not address 'OpenID Connect user profile field mapping'. The third node provides 'general information about the OpenID Connect Authenticator extension' but lacks details on 'mapping user profile fields'. The fourth node lists 'dependencies for the OpenID Connect Authenticator extension', which is unrelated to 'user profile field customization'. Finally, the fifth node mentions 'configuration guides for specific systems' but does not provide information on 'customizing user profile fields with OpenID Connect'.

ContextualRecall: The score is 0.40 because while the retrieval context supports the general idea of customizing user profile fields using 'xwiki.properties' and OpenID Connect (as seen in the 1st node in retrieval context), it lacks specific examples and detailed steps mentioned in the expected output, such as using 'oidc.user.mapping', replacing property names, and the need to save and restart XWiki.

CustomContextualRelevancy: The retrieval context does not mention customizing user profile fields or using 'oidc.user.mapping' in 'xwiki.properties', nor does it provide examples or instructions related to these mappings.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6128205128205129

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.23076923076923078

CustomContextualRelevancy: 0.5

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. 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, directly addresses group synchronization by mentioning the 'oidc.groups.mapping' property and the need to define a mapping between XWiki groups and provider groups. However, irrelevant nodes, such as the first node discussing authentication management and OpenID Connect provider

endpoints without mentioning group synchronization, and the second node about container authentication, are ranked higher than the relevant node.

ContextualRecall: The score is 0.23 because while the 3rd node in retrieval context supports configuring group synchronization in XWiki using 'oidc.groups.claim' and 'oidc.groups.mapping', many specific details such as locating and editing the 'xwiki.properties' file, setting and replacing specific properties, and restarting XWiki are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using oidc.groups.mapping but lacks details on configuring oidc.groups.claim, 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: 0.3333333333333333

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.67 because the output contains relevant information addressing the file upload restrictions, but it is lowered due to multiple irrelevant source link statements that do not directly contribute to solving 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 to the input. The first node discusses 'language preferences and translations in XWiki', which are unrelated to file upload restrictions. The second node describes 'various XWiki applications and features', but does not address file upload restrictions. The third node is about 'CSS and XSL for PDF styling in XWiki', which does not relate to configuring file upload restrictions. The fourth node lists 'various XWiki macros and features', but does not mention file upload settings. The fifth node contains 'image links and references to various projects', which are not relevant to file upload configurations in XWiki. The sixth node provides 'image links and references to different projects', which do not pertain to file upload restrictions.

ContextualRecall: The score is 0.00 because none of the instructions or configurations mentioned in the expected output can be attributed to any node(s) in the retrieval context. There is no information in the retrieval context about MIME type configuration, file size limits, or specific steps for editing the XWikiPreferences.

CustomContextualRelevancy: The retrieval context does not contain any information about configuring MIME types or setting maximum attachment size.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.6938461538461539

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.77 because the response mostly addresses the question about refusing script rights to a space administrator, 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. Keep up the great work!

ContextualPrecision: The score is 0.50 because the second node in the retrieval context, which is relevant as it 'discusses programming and script rights, mentioning that script rights can be denied,' is ranked after the first node. The first node is irrelevant since it 'is about language preferences and translations in XWiki, which is unrelated to rights management.' Additionally, other irrelevant nodes, such as the third node, 'lists various rights and their default states but does not specifically address the inability to deny script rights to an administrator,' are ranked higher than the relevant node.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the retrieval context, highlighting that administrative rights cannot be denied, as supported by the relevant node. Great job!

CustomContextualRelevancy: The retrieval context mentions that certain rights like ADMIN are not deniable, but it lacks specific information about the inability to refuse rights implied by admin rights, as stated in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.34285714285714286

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about configuring XWiki permissions, but includes repeated irrelevant statements about 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 'the structure of an XWiki page in XML format', which does not relate to configuring access rights for teams. The second node talks about 'sending representations and modifying resources using XML', which is not relevant to setting access permissions in XWiki. The third node is about 'editing comments and attachments', which is unrelated to configuring access rights for team spaces. The fourth node explains 'the emulation of spaces and sub-pages in XWiki', but does not provide information on setting access rights for teams. The fifth node provides 'a list of actions related to migrating and renaming pages', which is not pertinent to configuring access rights for teams. Finally, the sixth node discusses 'configuration display settings in XWiki', which does not address the question of 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, defining rights at wiki or space levels, or inheriting rights for child pages as mentioned in the expected output.
Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7301298701298702
Individual_scores:
AnswerRelevancy: 0.6363636363636364
Faithfulness: 1.0
ContextualPrecision: 0.5
ContextualRecall: 0.7142857142857143
CustomContextualRelevancy: 0.8
Reasons:

AnswerRelevancy: The score is 0.64 because the output partially addresses the question about authenticating users with access tokens, but the repeated irrelevant statements about source links lower 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 0.50 because relevant nodes in the retrieval context are not consistently ranked higher than irrelevant nodes. The first node, ranked 1, is irrelevant as it 'does not mention token-based authentication or JWTs.' The third node, ranked 3, is also irrelevant since it 'is not directly related to the use of JWTs for authentication.' Additionally, the fifth node, ranked 5, 'does not specifically address token-based authentication or JWTs.' However, the second node, ranked 2, is relevant as it 'discusses the configuration of 'Authorized Applications' which is relevant to setting up token-based authentication,' and the fourth node, ranked 4, 'provides a description of the extension that allows authentication using JWTs,' supporting the expected output.

ContextualRecall: The score is 0.71 because while the retrieval context nodes provide information about authenticating users with JWTs, enabling the token authenticator, and configuring authorized applications, they lack details on generating a JWT token with required claims and signing key requirements, and including the token in the 'Authorization' header.

CustomContextualRelevancy: The retrieval context includes details about configuring authorized applications and enabling the token authenticator, matching most of the expected answer, but lacks specific mention of JWT claims and Ed25519 key.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_032_result.json

Overall_score: 0.36363636363636365
Individual_scores:
AnswerRelevancy: 0.8181818181818182
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.82 because the main response addresses the question about authenticating users with access tokens effectively. However, it includes unrelated source links that slightly detract from its 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 'model validation strategies and configuration in XWiki', which is unrelated to 'authenticating users with access tokens'. The second node is about 'attachment validation and configuration in XWiki', which does not relate to 'authenticating users with 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 linked 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, JWT, or the LLM-Anwendung (BETA)-Erweiterung as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.36363636363636365

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the question about authenticating users with access tokens, but it includes unrelated 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 'upgrading XWiki' without mentioning 'user authentication with access tokens'. The second node talks about 'features of the Administration Application' but lacks any reference to 'authentication with access tokens'. The third node focuses on 'managing user rights and profiles' and, although it mentions an 'Authentication guide', it does not directly address '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.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication, JWT, or the specific configuration steps mentioned in the expected output.

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: 0.4764705882352941

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the response is mostly relevant and helpful, addressing the issue of the missing bell icon in XWiki. However, it includes repeated source links that do not directly address the specific problem, which prevents the score 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 0.50 because the second node in the retrieval context is relevant, mentioning the 'Alert menu' and the bell icon, which directly relates to the user's issue. However, the first node discusses 'installation instructions and prerequisites for extensions', which is not directly related to the problem, leading to a lower score. Additionally, the third node, which 'lists dependencies for the notifications extension', and the fourth node, which 'talks about overriding notification templates', are also irrelevant to the issue of enabling notifications or fixing the bell icon, and should be ranked lower than the relevant node.

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

CustomContextualRelevancy: The retrieval context does not mention enabling notifications via the 'notifications.enabled' setting in 'xwiki.properties', which is the core information in the expected answer.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.75

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.75

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

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 0.75 because the relevant nodes in the retrieval context are mostly ranked higher than irrelevant nodes. The first node is relevant as it 'explains how to follow a user and receive notifications about their activities,' which directly addresses the issue. However, the second node, which 'discusses filters for notifications,' and the third node, which is about 'getting an RSS feed for notifications,' are not directly relevant and should be ranked lower. Additionally, the fifth node, which discusses the 'What's New' application, is also not directly relevant to the problem of not receiving notifications and should be ranked lower than the relevant nodes.

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 on the 'alert' menu (node 4) and user following (node 1). However, key instructions like enabling notifications in `xwiki.properties` and subscribing to pages or users are not found in the nodes.

CustomContextualRelevancy: The retrieval context mentions following users and watching pages, but lacks details on enabling notifications in `xwiki.properties` and using the alert menu.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.9733333333333334

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.87 because the response is mostly relevant and helpful, but it includes repeated irrelevant statements about a source link that do not directly address how to enable notifications for one's own actions in XWiki.

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 explains that 'by default, one don't receive notification about one's own activity' and mentions the ability to 'disable this filter in the "Advanced filtering options" section of your own notification settings,' which directly addresses 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,' which is relevant to the expected output. Irrelevant nodes, such as the third node discussing administrators configuring notifications, the fourth node about RSS feeds, and the fifth node providing general notification methods, are ranked lower, ensuring a perfect score. Great job!

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 fully contains 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 user's notification settings.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5394117647058823

Individual_scores:

AnswerRelevancy: 0.6470588235294118

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.65 because the output partially addresses the question about disabling notifications in XWiki, but includes several irrelevant details about types of notifications without directly explaining how to disable them.

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,' which doesn't address disabling automatic notifications. The second node talks about 'following a user and default watch settings,' but lacks information on disabling notifications. The third node covers 'filtering notifications,' without mentioning how to disable automatic notifications. The fourth node is about 'enabling/disabling email notifications,' not related to the input query. The fifth node discusses 'administrative settings for notifications,' but doesn't help users disable automatic notifications. Lastly, the sixth node is about 'overriding default notification templates,' which is not relevant to the input.

ContextualRecall: The score is 0.75 because while nodes in the retrieval context support the idea of automatic page watching and notification settings (sentence 1 and 3), they do not mention being notified about changes on your work (sentence 2).

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8433879781420766

Individual_scores:

AnswerRelevancy: 0.9836065573770492

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.98 because the response is highly relevant and addresses the OpenID Connect configuration issue effectively, with only a minor irrelevant mention of LDAP Authenticator, which 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 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 most pertinent information is prioritized, making the retrieval highly effective. Great job on getting it spot on!

ContextualRecall: The score is 0.33 because while the 1st node in the retrieval context supports the information about the 'oidc.enableUser=true' property, it lacks details on adding or updating properties, uncommenting the property, restarting XWiki, and the automatic enabling of user accounts on first login.

CustomContextualRelevancy: The retrieval context contains the key information about uncommenting and setting the 'oidc.enableUser' property to 'true' in the xwiki.properties file, which matches the expected output. However, it lacks explicit instructions about saving the file and restarting the XWiki instance.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.9088888888888889

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant and effectively addresses the question about customizing user profile fields in XWiki. However, it includes a source link that does not directly contribute to the solution, 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 mentions 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user' and provides examples of mapping custom properties, is ranked first. This directly relates to customizing user profile fields. The irrelevant nodes, such as the second node discussing 'requesting access to various claims at the UserInfo endpoint' and not about customizing user profile fields, are ranked lower, maintaining perfect contextual precision. Great job!

ContextualRecall: The score is 0.60 because, while the initial part about customizing user profile fields and mapping properties is supported by the 1st node in the retrieval context, the instructions on replacing properties and restarting XWiki are not directly supported by any node in the retrieval context.

CustomContextualRelevancy: The retrieval context contains the exact information found in the expected answer, including the use of the `oidc.user.mapping` property in the `xwiki.properties` file to map user profile fields from the OpenID Connect provider.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.889090909090909

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5454545454545454

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 1.00 because the output is completely 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 crucial information about `oidc.groups.claim` and `oidc.groups.mapping`, which are essential for configuring group

synchronization, while the fourth and fifth nodes only offer general OpenID Connect information and authentication management, which are not directly related to the group synchronization process.

ContextualRecall: The score is 0.55 because while nodes in retrieval context provide specific examples and details for configuring properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, they lack coverage of introductory and generic instructional content present in the expected output.

CustomContextualRelevancy: The retrieval context covers configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', matching the expected answer. However, it lacks explicit instructions on saving the file and restarting XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.42000000000000004

Individual_scores:

AnswerRelevancy: 0.06666666666666667

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.07 because the output contains numerous irrelevant statements related to XWiki and MediaWiki configurations, which do not directly address the input question about setting upload limits on a different wiki platform. The presence of these unrelated details significantly lowers the relevancy 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 0.33 because the relevant node in the retrieval context, which 'explains how to set the maximum size of an attachment in XWikiPreferences', is ranked third. However, it should be ranked higher than the irrelevant nodes. For instance, the first node 'discusses linking to external files and images, backlinks indexation, and upload summaries', which is not related to configuring allowed mimetypes or setting maximum attachment sizes. Similarly, the second node 'provides information about XWiki's structure and page features', which does not relate to configuring attachment upload settings or size limits.

ContextualRecall: The score is 0.50 because while the 3rd node in the retrieval context supports the steps related to configuring the maximum attachment size (Steps 2.1 to 2.5), there is no information in the retrieval context about configuring allowed mimetypes or editing the xwiki.properties file as described in Step 1 of the expected output.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific instructions for 10MB limit and allowed mimetypes configuration.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.7684210526315789

Individual_scores:

AnswerRelevancy: 0.8421052631578947

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.84 because the response largely addresses the question of denying script rights to a space administrator, but it includes repeated 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. 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 '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 settings' but does not mention the inability to deny script rights to a space administrator. 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 'rights conversion and exclusion of pages and spaces,' which 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 nodes in retrieval context. Great job!

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7511904761904762

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.875

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.71 because the response partially addresses the question about configuring spaces in XWiki by providing some relevant information. However, it includes several irrelevant statements related to source links and topics not directly connected to the specific configuration question, which 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 0.67 because relevant nodes in the retrieval context are ranked higher than some irrelevant nodes. The first node discusses 'rights inheritance and administration features,' which is relevant to configuring permissions in XWiki. However, the second node, which is about 'parent/child relationship and the introduction of Nested Pages,' and the third node, which focuses on 'string formatting and OpenID Connect properties,' are not directly related to the input question and should be ranked lower. The fourth node provides information on 'XWiki's permission system,' which is directly relevant, and the sixth node discusses 'setting wiki-wide and page-level rights,' which is pertinent to the input. The presence of irrelevant nodes ranked higher than some relevant ones prevents a higher score.

ContextualRecall: The score is 0.88 because most of the expected output is supported by the nodes in the retrieval context. Specifically, nodes in the retrieval context provide information on creating groups, setting wiki-wide and space-level rights, and inheriting rights for child pages, which correspond to sentences 1 through 4 in the expected output. However, there is no mention of adjusting rights for additional teams or spaces in the retrieval context, which affects the completeness of the information provided.

CustomContextualRelevancy: The retrieval context mentions setting rights at wiki and space levels and rights inheritance, but lacks details on creating groups for each team and specific steps for setting 'View' and 'Edit' rights as outlined in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.5333333333333333

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

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 0.17 because the relevant node in the retrieval context, which discusses editing the `xwiki.properties` file to enable notifications, is ranked last (6th node). The preceding nodes are irrelevant as they discuss topics like 'custom notification filters', 'alert menu representation', 'various ways to receive notifications', 'links to notification settings', and 'customizing notification settings', 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 supports the idea of editing the xwiki.properties file, it does not specifically mention setting 'notifications.enabled' to 'true', leading to partial alignment with the expected output.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3333333333333333

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.67 because the output partially addresses the issue of not receiving notifications, but includes an irrelevant source link that does 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. The first node includes 'image links and references which do not relate to enabling notifications or following pages/users in XWiki.' Similarly, the second node 'consists of image links and references, which are unrelated to the process of enabling notifications or following entities in XWiki.' The third node also 'provides image links and references, offering no relevant information about managing notifications or following pages/users in XWiki.' As a result, none of the nodes provide relevant information for the input.

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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.36363636363636365

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.82 because the answer largely 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 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 notification settings. The second node describes 'an application for administrating wiki features', which does not pertain to user notification settings. The third node includes information about 'various extensions like Office Importer and Syntax Highlighting', none of which relate to enabling notifications for one's own activity. 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 does not contain any information about notification settings or user activity filters mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.32857142857142857

Individual_scores:

AnswerRelevancy: 0.6428571428571429

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.64 because the output partially addresses the question about disabling notifications, but includes several irrelevant links to documentation pages that do not pertain to the issue at hand.

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 and managing attachments', which is unrelated to managing notifications. The second node is about 'moving and deleting attachments', which does not relate to notification settings. The third node provides information on 'displaying content and macros', which is not relevant to disabling notifications. The fourth node is about 'administrating wiki features like users and groups', which does not pertain to notification settings. The fifth node discusses 'permissions and security modules', which are unrelated to managing notifications. Lastly, the sixth node lists 'various extensions and applications', none of which relate to disabling 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 contain any information related to automatic page following, notification settings, or 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: 0.52

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.0

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 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 activating user accounts on first login with OpenID Connect. The second node provides 'general information about OpenID Connect authentication and its installation', lacking the specific configuration details needed. The third node talks about 'OpenID Connect authenticator and its compatibility' without addressing user account activation. Lastly, the fourth node lists 'various OpenID Connect endpoints and properties' but fails to mention the specific property 'oidc.enableUser' or any instructions for automatic activation.

ContextualRecall: The score is 0.60 because while the retrieval context (specifically node 4) supports the instructions about adjusting the 'xwiki.properties' file and setting the 'oidc.enableUser' property, it

lacks explicit details on saving the file, restarting the XWiki instance, and ensuring automatic activation of new user accounts upon first login via OpenID Connect.

CustomContextualRelevancy: The retrieval context does not mention the property 'oidc.enableUser' or the steps to update 'xwiki.properties' for automatic user account activation.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.46947368421052627

Individual_scores:

AnswerRelevancy: 0.9473684210526315

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant and addresses the customization of user profile fields from an OpenID Connect provider to XWiki user properties.

However, it slightly deviates by mentioning interface customization related to Google, which is not directly pertinent 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. The first node discusses 'authentication management and OpenID Connect endpoints', the second node is about 'container authentication and other authentication methods like LDAP', the third node provides 'OpenID Connect authentication and configuration guides', the fourth node describes 'the OpenID Connect Authenticator extension, its developer, and compatibility', and the fifth node lists 'dependencies for the OpenID Connect Authenticator extension'. None of these nodes address mapping user profile fields or the 'xwiki.properties' file, which is the focus of the input.

ContextualRecall: The score is 0.40 because while nodes in the retrieval context mention 'xwiki.properties' and 'oidc.user.mapping', which support the customization of user profile fields, they lack details on the example mapping syntax, replacement instructions, and the steps to save and restart XWiki.

CustomContextualRelevancy: The retrieval context lacks details on customizing user profile fields in xwiki.properties using oidc.user.mapping, as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.4

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

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 provider configuration, but does not mention group synchronization or the specific properties needed for it.' The second node 'focuses on container authentication and other authentication methods like LDAP, without addressing group synchronization with OpenID Connect.' The third node 'mentions OpenID Connect authentication and provides links to configuration guides, it does not cover the specific steps or properties needed for group synchronization.' The fourth node 'lists dependencies for the OpenID Connect authenticator but does not provide any information relevant to group synchronization setup.' Finally, the fifth node 'provides general information about the OpenID Connect authenticator, including compatibility and installation, but does not mention group synchronization.'

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 all instructions and examples provided are not covered or mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context lacks any information about configuring group synchronization, properties like `oidc.groups.claim`, `oidc.groups.separator`, or `oidc.groups.mapping`, which are detailed in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.60453216374269

Individual_scores:

AnswerRelevancy: 0.8421052631578947

Faithfulness: 1.0

ContextualPrecision: 0.325

ContextualRecall: 0.555555555555556

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.84 because the response mostly addresses the question about configuring upload restrictions for images and PDFs, but includes irrelevant links that do not directly relate to the topic, slightly lowering 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 0.33 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 'features and macros available in XWiki', which do not relate to configuring attachment upload restrictions. These should be ranked lower than the fourth and fifth nodes, which are directly relevant as they include links to 'Size Limit' and 'Mimetype Restriction' and provide steps on setting the maximum size of an attachment.

ContextualRecall: The score is 0.56 because while the retrieval context's 5th node supports the steps for configuring the maximum attachment size (related to sentences in 'Schritt 2'), it lacks information on configuring MIME types and restricting uploads to images and PDFs (related to sentences in 'Schritt 1').

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 the exact steps for setting the size to 10 MB.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_019_result.json

Overall_score: 0.36923076923076925
Individual_scores:
AnswerRelevancy: 0.8461538461538461
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.85 because the answer mostly addresses the question about denying script rights to a domain administrator, 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'form validation and JavaScript', which is unrelated to script rights or administrative permissions in XWiki. The second node is about 'template providers and page creation in XWiki', which also does not relate to script rights. The third node describes 'a custom right check class and configuration in XWiki', but does not address the issue of denying script rights to an area administrator. Since all nodes are irrelevant, they should be ranked lower, resulting in a score of 0.00.

ContextualRecall: The score is 0.00 because the expected output sentence about admin rights in XWiki is not supported by any information in the nodes in retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to the admin rights or their implications as mentioned in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_020_result.json

Overall_score: 0.3544117647058823
Individual_scores:
AnswerRelevancy: 0.6470588235294118
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.125
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.65 because the output contains multiple irrelevant source links and a misleading claim about access control in XWiki, which detracts from fully addressing the input question. However, it still provides some relevant information about configuring access in XWiki, which is why it 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 query. The first node discusses 'XWiki's database efficiency and programming features', which do not pertain to configuring access controls for team areas. The second node is about 'notifications and

application details', unrelated to access permissions. The third node explains 'nested pages and subwikis', not directly related to team-specific access rights. The fourth node provides information on 'managing users and groups' but does not specifically address setting access rights for team areas. Lastly, the fifth node is about 'group synchronization and customization', which does not directly relate to configuring access rights for team areas in XWiki.

ContextualRecall: The score is 0.12 because only one supportive reason was identified, linking sentence 1 to node 4 in the retrieval context, while the rest of the expected output lacks corresponding nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about access control steps or rights management mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.4666666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the response mostly addresses the issue with the unavailable bell icon, 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', the second node lists 'various notification-related options and links', the third node describes 'different notification preferences and macros', the fourth node provides a description of 'the notifications application and its features', and the fifth node discusses 'RSS feeds and notification settings'. None of these address 'enabling notifications in the wiki settings', which is relevant to the input.

ContextualRecall: The score is 0.50 because the 1st sentence of the expected output is supported by the 1st node in the retrieval context mentioning 'notifications' and 'wiki'. However, the 2nd sentence lacks support as none of the nodes in the retrieval context mention 'defining the property `notifications.enabled` in `xwiki.properties` on `true`'.

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.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.7567032967032967

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.77 because the output contains some relevant information addressing the notification issue, but it is diluted by multiple irrelevant statements such as source links, incomplete information, and general labels that do not directly 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 in the retrieval context are ranked higher than irrelevant nodes. The first and second nodes relate to 'ensuring notifications are activated' and 'guidance on following users or pages for notifications', which are directly relevant to the input. The third node, which focuses on 'Get a notification RSS feed' and 'Notifications Macro', and the fourth node, listing 'various notification preferences macros', are not directly related to enabling notifications or following users/pages, and are correctly ranked lower.

ContextualRecall: The score is 0.71 because several key aspects of the expected output, such as enabling notifications and following users, are supported by nodes in the retrieval context. However, specific details like 'xwiki.properties' and 'profil utilisateur' are missing, affecting the overall alignment.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks specific details about setting 'notifications.enabled' to 'true' in 'xwiki.properties' and using the alert menu for notifications.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8244444444444445

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 0.8333333333333334

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.89 because the response effectively addresses how to enable notifications for one's own actions, but it includes an irrelevant link to XWiki syntax documentation, which slightly detracts from its relevance.

Faithfulness: The score is 0.83 because the actual output incorrectly claims that you can enable notifications for your own actions within an app, while the retrieval context clarifies that by default, you won't receive notifications for actions done by yourself.

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, directly addressing the input question about receiving notifications for one's own actions. 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 about disabling the filter in 'Options de filtrage avancées' is not found in any of the 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 in settings, aligning closely with the expected output. However, it does not explicitly mention the 'Options de filtrage avancées' section.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.9236363636363636

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the question about disabling notifications on Facebook, but it includes irrelevant source links that do not pertain to the topic, slightly reducing its relevancy.

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 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 'Auto Watch', which are directly related to disabling automatic notifications. The irrelevant node, ranked fourth, discusses 'Get a notification RSS feed' and 'Notifications Macro', which are not related to the input 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 includes information about auto-watched pages and custom notification filters, aligning with the expected answer's details on automatic page following and filter management. However, it lacks explicit mention of modifying these settings in the user's notification profile.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.4257142857142857

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and provides useful information about configuring OpenID Connect in XWiki, but it includes an irrelevant mention of LDAP Authenticator, which slightly detracts from the focus on the specific question asked.

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 contexts are irrelevant to the input. The first node discusses 'domain-based authentication and session cookie configuration', which does not relate to enabling user accounts automatically via OpenID Connect. The second node is about 'container authentication and LDAP', lacking the specific property needed for OpenID Connect. The

third node mentions OpenID Connect but misses the specific configuration for automatic activation. The fourth node lists dependencies for the OpenID Connect extension without addressing the needed configuration. The fifth node focuses on 'LDAP and Active Directory integration', unrelated to the OpenID Connect configuration required.

ContextualRecall: The score is 0.20 because only the mention of the property `oidc.enableUser` in sentence 1 is somewhat related to node 1 in the retrieval context, while the rest of the expected output lacks support from the nodes, including key instructions and guarantees.

CustomContextualRelevancy: The retrieval context does not contain information about adjusting the xwiki.properties file or the specific property oidc.enableUser for automatic activation of user accounts on first login using OpenID Connect.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.30666666666666664

Individual_scores:

AnswerRelevancy: 0.3333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.33 because the output contains multiple irrelevant details about directory paths, .jar files, and technical instructions that do not directly address the question about mapping additional user profile fields from an OpenID Connect provider. However, the presence of some relevant information prevents the score from being 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 'authentication management class and OpenID Connect provider endpoints', but does not mention 'user profile field mapping or customization'. The second node explains 'various authentication methods like Container Authentication and Active Directory', which does not address 'mapping user profile fields from OpenID Connect'. The third node provides information about the 'OpenID Connect Authenticator extension', but does not discuss 'mapping user profile fields or customizing them'. The fourth node lists 'dependencies for the OpenID Connect Authenticator extension', which is unrelated to 'mapping user profile fields'. The fifth node provides 'links to configuration guides and describes the OpenID Connect Authenticator', but does not mention 'user profile field mapping'.

ContextualRecall: The score is 0.20 because only the first sentence in the expected output is supported by the 1st node in the retrieval context, which mentions 'xwiki.properties' and mapping fields from OpenID Connect provider. However, the remaining sentences lack support as none of the nodes in the retrieval context mention 'oidc.user.mapping', provide examples of mapping custom properties, or discuss the steps to apply changes.

CustomContextualRelevancy: The retrieval context does not contain any information about customizing user profile fields using `oidc.user.mapping` in `xwiki.properties`, which is the main focus of the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6155555555555555

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.4444444444444444

CustomContextualRelevancy: 0.3

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. Everything aligns perfectly with the retrieval context. Great job!

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, ranked third, mentions 'group synchronization and the use of the `oidc.groups.mapping` property', which is directly related to the input query about configuring group synchronization. However, irrelevant nodes ranked first, second, fourth, and fifth discuss 'authentication management and OpenID Connect endpoints', 'container authentication and other methods', 'general information about OpenID Connect authentication', and 'dependencies for the OpenID Connect extension', none of which address group synchronization, causing the score to be lower.

ContextualRecall: The score is 0.44 because while the 3rd node in the retrieval context provides relevant information on configuring group synchronization and properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', it lacks specific details such as locating and saving the 'xwiki.properties' file and replacing placeholder values with actual ones, which are present in sentences 1, 3, 5, and 7 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 as outlined in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.35

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.75 because the response mostly addresses the question about file upload restrictions, but it includes multiple irrelevant source links that do not directly relate to 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 discusses 'language preferences and translations in XWiki', which does not relate to configuring file upload restrictions. The second node talks about 'various administrative applications and features in XWiki', but again, it does not address file upload settings. The third node is about 'PDF

export customization using CSS and XSL', which is not relevant to the topic of file upload restrictions. The fourth node describes 'various macros and their functionalities in XWiki', none of which pertain to file upload restrictions. The fifth node contains 'links to images and references', offering no information about file upload settings. Lastly, the sixth node is similar to the fifth, with 'links to images and references', unrelated to file upload configurations.

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 configuring MIME types or attachment size limits as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.6971428571428572

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.79 because the main content addresses the question about refusing 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 0.50 because the second node in the retrieval context, which is relevant as it 'discusses programming and script rights, indicating that script rights are part of the admin rights and cannot be denied separately,' is ranked lower than the first node, which is irrelevant since it is 'about language preferences and translations in XWiki, which is unrelated to script rights or administrative permissions.' Additionally, other irrelevant nodes such as the third node, which 'lists various rights and their default states but does not specifically address the inability to deny script rights to an admin,' are ranked higher than the relevant node.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the information in the retrieval context. Great job!

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.35

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.75 because the response addresses the main question about configuring spaces 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 provides 'XML metadata about a page, which is unrelated to configuring access permissions in XWiki.' The second node discusses 'sending representations to modify resources, which does not relate to access control settings for spaces in XWiki.' The third node is about 'managing comments, which is unrelated to setting up team-specific access permissions in XWiki.' The fourth node describes 'how pages and spaces are created in XWiki, but does not provide information on configuring access rights for teams.' The fifth node lists 'actions needed after running a migrator, which is not relevant to configuring access permissions in XWiki.' Lastly, the sixth node discusses 'configuration display options in XWiki, which does not directly relate to setting up 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, indicating a complete lack of relevant information in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to setting access rights or managing teams in XWiki as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.6828571428571429

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.7

Reasons:

AnswerRelevancy: The score is 0.50 because while the output includes source links that could potentially be useful, it fails to directly address the question about authenticating users with access tokens, leading to a lack of direct 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 some relevant nodes are ranked lower than irrelevant nodes. The first node is about 'LDAP configuration' and does not mention token-based authentication, which should be ranked lower. Similarly, the third node discusses 'OpenID Connect scopes and response types', unrelated to token-based authentication, and should also be ranked lower. The fifth node gives a general overview of 'XWiki's authentication mechanisms' without addressing token-based authentication, and should be ranked lower as well. However, the second and fourth nodes are correctly identified as relevant, discussing 'Authorized Applications Configuration' and 'authentication with JWT', respectively.

ContextualRecall: The score is 0.71 because while the retrieval context nodes provide substantial information about authenticating users with JWTs, enabling the token authenticator, and configuring authorized applications, they lack specific details about generating a JWT token with required claims

and including the token in the 'Authorization' header.

CustomContextualRelevancy: The retrieval context mentions using JWT for authentication and configuring authorized applications, aligning with the expected answer. However, details like specific claims in the token and the 'Authorization' header format are missing.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.37777777777777777

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.89 because the answer is mostly relevant to the question about authenticating users with access tokens, but it includes a source link that is not directly relevant, 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', the second node talks about 'attachment validation and configuration', and the third node provides details on 'the admin UI for model validation'. None of these are related to 'authenticating users 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, as the context does not cover topics like 'Benutzer mit Zugriffstoken', 'Token-basierten Authentifizierung', 'externen Anwendungen', 'JSON Web Tokens (JWT)', or any other related concepts.

CustomContextualRelevancy: The retrieval context lacks any information related to token-based authentication, JWT, or the LLM-Anwendung (BETA)-Erweiterung mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.34545454545454546

Individual_scores:

AnswerRelevancy: 0.7272727272727273

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.73 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. 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 lacks information on 'authenticating users with access tokens'. The second node discusses 'features related to users and rights' but does not cover 'authenticating users with access tokens'. The third node mentions 'managing user rights and profiles' without addressing '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, indicating a complete mismatch between the expected content and the provided context.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication, JWT, or related configurations mentioned in the expected output.

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: 0.39

Individual_scores:

AnswerRelevancy: 0.95

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant and effectively addresses the issue of the missing bell icon in XWiki. However, it includes source links that do not directly contribute to solving the problem, 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. The first node discusses 'installation instructions and prerequisites for extensions', which does not relate to enabling notifications. The second node describes 'the functionality and appearance of the notifications menu', but lacks information on enabling notifications. The third node lists 'dependencies for the notifications extension', not addressing the issue of enabling notifications. Lastly, the fourth node talks about 'overriding notification templates', which is unrelated to the problem of 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` file.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications or the `notifications.enabled` setting in `xwiki.properties`, which are key elements in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.5846991509782207

Individual_scores:

AnswerRelevancy: 0.046511627906976744

Faithfulness: 1

ContextualPrecision: 0.8055555555555555

ContextualRecall: 0.5714285714285714

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.05 because the output contains numerous irrelevant and nonsensical statements that do not address the issue of not receiving notifications in XWiki. The low score reflects the lack of relevant content, but it is not zero because there might be a minimal connection or attempt to address 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 0.81 because most relevant nodes are ranked higher, providing useful information on following users and accessing notification settings in XWiki. However, the second node, discussing 'filtering notifications and hidden pages', and the fifth node, about the 'What's New' application, are ranked higher than they should be as they do not directly address enabling notifications.

ContextualRecall: The score is 0.57 because while there are several supported sentences, such as those related to using the 'alert' menu and following users (supported by nodes in the retrieval context), key instructions about enabling notifications in `xwiki.properties` and subscribing to pages or users are not found in the nodes.

CustomContextualRelevancy: The retrieval context covers following users and notification settings but lacks details on enabling notifications in xwiki.properties and watching pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.990909090909091

Individual_scores:

AnswerRelevancy: 0.9545454545454546

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant to the input, addressing the issue of enabling notifications for one's own actions in XWiki. However, it includes vague suggestions like 'Check User Permissions' and 'Review Application Settings' that do not directly solve the problem, preventing a perfect score.

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 are ranked higher than 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.' This directly addresses the input question. The second node mentions: 'By default, you won't receive notifications for actions done by yourself. This can be changed by switching off the Own event filter.' This is relevant to the expected output. Irrelevant nodes, such as the third node discussing administrators' capabilities, the fourth node focusing on RSS feeds, and the fifth node providing general notification methods, are all 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 change this setting 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: 0.6160130718954248

Individual_scores:

AnswerRelevancy: 0.9411764705882353

Faithfulness: 0.8888888888888888

ContextualPrecision: 0.2

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and provides useful information on disabling notifications in XWiki. However, it includes a source link that doesn't directly address the question, preventing a perfect score.

Faithfulness: The score is 0.89 because the actual output incorrectly claims that 'New Page Alert' might be automatically enabled by default for all users, whereas the retrieval context clarifies that in XWiki 15.5+, by default, nothing is watched for new users.

ContextualPrecision: The score is 0.20 because the relevant node in the retrieval context, ranked fifth, discusses the 'default auto-watched mode' and user settings for disabling notifications, which is pertinent to the input query. However, the first four nodes are irrelevant as they focus on notification RSS feeds, following users, filtering notifications, and email notifications, none of which address disabling notifications or modifying automatic page watching settings. This misalignment in ranking significantly lowers the score.

ContextualRecall: The score is 0.75 because while nodes in the retrieval context support the expected output regarding default settings, user profile adjustments, and filter management (sentences 1, 3, and 4), there is no mention of notifications about changes on your work (sentence 2).

CustomContextualRelevancy: The retrieval context mentions default notification settings and user profile settings, but lacks details on automatic page watching, notification filters, and specific settings adjustments found in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8523076923076923

Individual_scores:

AnswerRelevancy: 0.9615384615384616

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant to configuring OpenID Connect in XWiki, with only a minor mention of LDAP Authenticator, which is not pertinent 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 1.00 because the relevant node in the retrieval context, which contains the line '# oidc.enableUser=true', is ranked first. This directly addresses the configuration needed to enable user accounts upon first login using OpenID Connect. The irrelevant nodes, which discuss topics like domain-based authentication, general OpenID Connect information, container authentication, and LDAP configuration, are all ranked lower, ensuring precise and relevant information is prioritized.

ContextualRecall: The score is 0.40 because while the 1st node in the retrieval context supports the need to define the user account enabling property, it lacks information on adding or updating properties, saving the file, restarting XWiki, and the automatic enabling of new user accounts on first login via OpenID Connect.

CustomContextualRelevancy: The retrieval context includes the property '# oidc.enableUser=true', which matches the expected answer's requirement to uncomment and set this property to true. However, it lacks explicit instructions to save the file and restart the XWiki instance.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.9400000000000001

Individual_scores:

AnswerRelevancy: 0.95

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant and directly addresses the question about customizing user profile fields in XWiki. However, it includes a source link, which is not necessary for answering 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 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', is ranked first. This directly addresses the input about customizing user profile fields. The irrelevant nodes discussing topics like 'access tokens and scopes' and 'OpenID Connect provider endpoints and HTTP headers' are correctly ranked lower, ensuring a perfect contextual precision score. Great job on getting everything in the right order!

ContextualRecall: The score is 0.75 because the retrieval context (node 3) supports the customization and mapping of user profile fields, including examples and instructions. However, it lacks information about saving the 'xwiki.properties' file and restarting XWiki, which is necessary to fully align with the expected output.

CustomContextualRelevancy: The retrieval context contains all the information found in 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.705

Individual_scores:

AnswerRelevancy: 0.125

Faithfulness: 1

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.12 because the output contains multiple nonsensical and incomplete statements that do not address the process of configuring group synchronization between XWiki and an OpenID Connect provider. The low score reflects the lack of relevant and coherent information provided in response 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 1.00 because all relevant nodes in the retrieval context are ranked higher than the irrelevant ones. The first three nodes provide detailed instructions on configuring group synchronization, such as setting the `oidc.groups.claim` property and configuring the `oidc.groups.mapping` property. The irrelevant nodes, ranked fourth and fifth, focus on general configuration and authentication management, which are not directly related to the group synchronization setup. Great job on ranking the relevant information at the top!

ContextualRecall: The score is 0.50 because while nodes in the retrieval context support the configuration steps for `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` properties (as seen in sentences 3, 5, and 7 of the expected output), they lack information on basic setup instructions like locating and editing the `xwiki.properties` file, replacing placeholders, and saving changes (as seen in sentences 1, 2, 4, 6, and 9).

CustomContextualRelevancy: The retrieval context contains detailed information on configuring the 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' properties as described in the expected output. It also mentions saving the 'xwiki.properties' file. However, it lacks explicit instructions on locating and opening the file with a text editor.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6072727272727273

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.6363636363636364

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.87 because the response mostly addresses the question about restricting uploads to images or PDFs with a size limit. However, it includes a general statement about XWiki concepts and mentions WordPress, which are not directly relevant to the specific 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 0.33 because the relevant node in the retrieval context, ranked third, provides information about setting the maximum size of an attachment in XWikiPreferences, which directly relates to configuring the maximum upload size. However, irrelevant nodes ranked first, second, fourth, and fifth discuss topics like linking to external files, XWiki's structure, attachment display, and versioning, which do not address configuring mimetypes or size limits for uploads. These irrelevant nodes should be ranked lower to improve the score.

ContextualRecall: The score is 0.64 because while the steps for configuring mimetypes and maximum attachment size are supported by nodes in the retrieval context, the overall instruction to limit uploads to images or PDFs of up to 10MB is not directly mentioned in the nodes.

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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.7823529411764706

Individual_scores:

AnswerRelevancy: 0.9117647058823529

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.91 because the response is largely relevant and provides useful information on denying script rights to a space administrator. However, the inclusion of unrelated source links slightly detracts from the 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 0.00 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 settings but does not mention the specific limitation of denying script rights to a space administrator. The fourth node provides code snippets related to 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 but does not relate to the denial of script rights to a space administrator.

ContextualRecall: The score is 1.00 because every aspect 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 right cannot be denied, which is confirmed by the retrieval context indicating that admin rights are not deniable and automatically imply other rights.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.7363333333333333

Individual_scores:

AnswerRelevancy: 0.84

Faithfulness: 1.0

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.875

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.84 because the answer largely addresses the question about configuring permissions in XWiki, but it includes multiple irrelevant source links that don't directly contribute to solving 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 0.67 because relevant nodes in the retrieval context, such as the first node discussing 'rights inheritance and administration features' and the fourth node providing information about XWiki's permission system, are ranked higher than some irrelevant nodes. However, the second node, focusing on 'the parent/child relationship and the introduction of Nested Pages', and the third node, discussing 'string formatting and OpenID Connect properties', are ranked higher than the fourth relevant node, which affects the score.

ContextualRecall: The score is 0.88 because most sentences in the expected output are well-supported by the nodes in the retrieval context. For instance, sentences 1 and 2 are supported by nodes 6 and 4 respectively, while sentence 4 is supported by node 1. However, the lack of information regarding adjusting rights for additional teams or spaces, as noted in the unsupportive reason, slightly impacts the score.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.512280701754386

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant to the issue of the unavailable bell icon, 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 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 that administrators can edit the xwiki.properties file to configure notification settings, is ranked sixth. This node should be ranked higher as it directly relates to enabling notifications. The first node discusses custom notification filters and their behavior, which is not related to enabling notifications in the wiki settings, and should be ranked lower. Similarly, the second node describes the notification bell and how notifications are displayed, but does not address enabling notifications, indicating it should also be ranked lower.

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 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`.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3777777777777777

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant to the issue of not receiving notifications, but it includes an irrelevant source link that 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 maintaining accuracy!

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 that do not relate to notifications or their settings.' The second node also consists of 'another set of image links and references, which do not provide any information about enabling notifications or related settings.' Similarly, the third node includes 'image links and references that are unrelated to the topic of notifications.' All nodes should be ranked lower as they do not address the input about notifications.

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 only contain image links and references.

CustomContextualRelevancy: The retrieval context contains image links and does not provide any information related to enabling notifications or following pages/users as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.376

Individual_scores:

AnswerRelevancy: 0.88

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the main content effectively addresses the question about enabling notifications for one's own activity, but the presence of source links and the unrelated word 'endlich' 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 other features unrelated to notification settings.' The second node is about 'administrating wiki features like users and groups, which does not relate to notification settings for user activity.' The third node covers 'importing office documents, sharing pages, and other features, but does not mention notification settings or user activity.' These irrelevant nodes are ranked higher than relevant ones, hence the low score.

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.

CustomContextualRelevancy: The retrieval context contains no information about notification settings or user activity filters as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.37435897435897436

Individual_scores:

AnswerRelevancy: 0.8717948717948718

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.87 because the main content effectively addresses how to disable notifications, but it is not higher due to multiple irrelevant source links that do not directly relate to the task of 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' which does not relate to notification settings. The second node is about 'moving attachments', unrelated to notifications. The third node lists 'various macros and their functions', none of which pertain to notifications. The fourth node describes an 'application for administrating wiki features', but does not mention notification settings. The fifth node is about 'controlling permissions and security', not related to notifications. The sixth node lists 'various XWiki extensions', none of which are 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 related to automatic page following or notification settings mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.48

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response perfectly addresses the question about configuring XWiki for automatic user activation with OpenID Connect, with no 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 'domain-based authentication and session cookies', which is unrelated to enabling user accounts upon first login with OpenID Connect. The second node provides 'general information about OpenID Connect authentication and installation instructions', but does not address the specific configuration needed. The third node describes 'compatibility and installation details', which do not pertain to user account activation configuration. The fourth node outlines 'various OpenID Connect endpoints and configurations', but does not mention the specific property 'oidc.enableUser' needed for automatic user activation.

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention the xwiki.properties file and OpenID Connect settings, they do not provide details on updating specific properties, restarting the XWiki instance, or automatic activation of user accounts on first login, which are key elements in the expected output.

CustomContextualRelevancy: The retrieval context does not contain information about automatically activating user accounts with OpenID Connect by modifying the xwiki.properties file.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.46571428571428575

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant to the input question about customizing user profile fields in XWiki. However, it includes source links that are not directly relevant to the specific query, preventing a perfect score.

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' without mentioning user profile field mapping. The second node talks about 'container authentication and other methods like Active Directory and LDAP', but not about mapping fields from OpenID Connect to XWiki. The third node provides 'links and descriptions about OpenID Connect authentication', yet lacks information on mapping additional user profile fields. The fourth node is a 'general description of the OpenID Connect Authenticator extension' without details on customizing user profile fields. Lastly, the fifth node lists 'dependencies for the OpenID Connect Authenticator extension' without discussing field mapping or using 'oidc.user.mapping'.

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention 'xwiki.properties' and 'oidc.user.mapping', which relate to sentences 1 and 2 of the expected output,

they lack specific examples and instructions found in sentences 3 and 4, such as replacing property names and restarting XWiki.

CustomContextualRelevancy: The retrieval context does not mention customizing user profile fields using `xwiki.properties` or `oidc.user.mapping`, which are key aspects of the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.23548387096774195

Individual_scores:

AnswerRelevancy: 0.1774193548387097

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.18 because the output contains numerous irrelevant statements that do not address the technical setup of group synchronization between XWiki and an OpenID Connect provider. The low score reflects the lack of relevant content, but it is not zero because there might be a minimal connection or attempt to address 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 maintaining accuracy!

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 mention group synchronization. The second node is about different authentication methods, lacking information on setting up group synchronization. The third node provides links to configuration guides for OpenID Connect authentication, but does not address group synchronization. The fourth node lists dependencies for the OpenID Connect authenticator extension, which is not relevant to group synchronization. Lastly, the fifth node provides general information about the OpenID Connect authenticator, but does not cover group synchronization setup.

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 lacks details about configuring group synchronization, properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, and restarting XWiki, which are essential in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.5923684210526317

Individual_scores:

AnswerRelevancy: 0.7368421052631579

Faithfulness: 1.0

ContextualPrecision: 0.325

ContextualRecall: 0.4

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.74 because the output provides relevant information on file type and size restrictions, but it is lowered due to multiple irrelevant source links that do not directly address the input 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 0.33 because the first three nodes in the retrieval context discuss topics such as 'trusted domains and PDF export customization', 'ODT and PDF export features', and 'various macros and features', which are unrelated to configuring attachment restrictions in a wiki. These irrelevant nodes are ranked higher than the fourth and fifth nodes, which include 'links to sections on Size Limit and Mimetype Restriction' and 'detailed instructions on setting the maximum size of an attachment and mentions MIME type restrictions', directly relevant to the input query. The presence of these relevant nodes, albeit ranked lower, contributes to the score being above zero.

ContextualRecall: The score is 0.40 because while some sentences in the expected output are supported by nodes in the retrieval context, such as editing the xwiki.properties file and setting the maximum attachment size, many critical details like the specific MIME type configuration and the steps to save changes are not found in the nodes, leading to a lower score.

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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.3826086956521739

Individual_scores:

AnswerRelevancy: 0.9130434782608695

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.91 because the response is highly relevant to the question about denying script rights to a domain administrator, but it includes source links 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 'form validation' and lacks any mention of 'script rights or XWiki admin rights.' The second node focuses on 'templates and page creation in XWiki,' which does not relate to 'script rights or admin rights.' The third node talks about 'a custom right check class and configuration,' but it does not address 'denying script rights to a space admin in XWiki.'

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 contain any information about XWiki not supporting the refusal of rights implied by the admin right.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.3722222222222223

Individual_scores:

AnswerRelevancy: 0.6111111111111112

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.25

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.61 because the response partially addresses the question about configuring access and editing permissions in XWiki. However, it includes multiple irrelevant statements, such as links to unrelated sources and general advice on team communication, which do not directly help in solving the specific configuration 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 XWiki's database support and programming features, which are not relevant to configuring access controls for team areas. The second node describes the Notifications Application in XWiki, which does not relate to setting access permissions for team areas. The third node explains the concept of Nested Pages and subwikis in XWiki, which does not directly address configuring access controls for team areas. The fourth node provides information on managing users and groups, but it does not specifically address setting access permissions for team areas. The fifth node discusses group synchronization and customization in XWiki, which is not directly related to configuring access permissions for team areas.

ContextualRecall: The score is 0.25 because only sentences 1 and 4 from the expected output are supported by nodes in the retrieval context, specifically nodes 4 and 3 respectively. The rest of the sentences are not supported by any nodes in the retrieval context, leading to a low score.

CustomContextualRelevancy: The retrieval context lacks any information about creating groups, setting wiki-wide or area-specific rights, or inheriting rights, which are all detailed in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.4666666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the response is mostly relevant to the issue of the notification bell not being available, but it includes links that are not directly related to solving 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'custom notification filters and their behavior', which does not relate to 'enabling notifications in the wiki settings'. Similarly, the second node lists 'various notification-related settings and options' but does not address the issue of enabling notifications. The third node describes 'different notification preferences and macros', again missing the focus on enabling notifications. The

fourth node provides a 'description of the notifications application and its menu' without mentioning enabling notifications. Lastly, the fifth node talks about 'obtaining an RSS feed for notifications and various settings', which is not relevant to enabling notifications in the wiki settings.

ContextualRecall: The score is 0.50 because while the retrieval context supports the general idea of activating notifications (related to sentence 1), it lacks specific details about configuring 'xwiki.properties' and setting 'notifications.enabled' to 'true' (related to sentence 2).

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.6228571428571429

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 1.0

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.00 because none of the statements in the output address the issue of not receiving notifications. All statements are unrelated to the input.

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

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node discusses 'Enable and disable the extension' and 'Enabling notifications for all users by default', which is directly related to ensuring notifications are activated. The second node covers 'Follow a user' and 'Watched Entities', which are pertinent to receiving notifications. Irrelevant nodes, such as the third node discussing 'RSS feeds and notification macros' and the fourth node focusing on 'displaying preferences and macros', are correctly ranked lower.

ContextualRecall: The score is 0.71 because while there are supportive connections between the retrieval context nodes and most of the expected output sentences, such as enabling notifications and following users, there are missing details like the specific mention of 'xwiki.properties' and the 'réseau' tab, which 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' in 'xwiki.properties' and specifics on following pages or users.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_023_result.json

Overall_score: 0.6799999999999999

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.00 because the output contains numerous irrelevant statements that do not address the query about enabling notifications for personal actions.

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, 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,' which is precisely what the input is about. The irrelevant nodes, which discuss topics like 'RSS feeds and general notification settings,' 'a list of links and sections related to notifications,' 'various notification preferences macros,' and 'technical aspects of notifications,' are all correctly ranked lower as they do not specifically address the issue of receiving notifications for one's own actions. Great job on getting it spot on!

ContextualRecall: The score is 0.50 because the first sentence is supported by the 1st node in the retrieval context, but 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, aligning closely with the expected output. However, it does not explicitly mention the 'Options de filtrage avancées' section.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_024_result.json

Overall_score: 0.9470967741935483

Individual_scores:

AnswerRelevancy: 0.9354838709677419

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant and provides a clear solution to disabling notifications, but it includes unnecessary links that do not directly address the question.

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 all relevant nodes in the retrieval context are ranked higher than the irrelevant node. The first three nodes discuss topics like 'Choosing the default auto-watched mode', 'Notifications Auto Watch Preferences', and 'Auto Watch' under 'Watched Entities', which are all pertinent to the user's query about managing and disabling automatic notifications. The irrelevant node, ranked fourth, discusses 'Get a notification RSS feed' and 'Notifications Macro', which does not directly address the user's issue. Great job on the ranking!

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 includes information about auto-watched pages, notification settings, and custom filters, aligning with the expected answer's details on automatic page following and notification settings. However, it lacks explicit mention of modifying the 'Suivi automatique des pages' setting directly in user profile notification settings.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.44000000000000006

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant with no 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 0.00 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 with OpenID Connect in XWiki. The second node is about 'container authentication' and mentions OpenID Connect, but does not address enabling user accounts automatically. The third node focuses on 'authentication providers and limitations', not on enabling user accounts automatically. The fourth node lists 'dependencies for the OpenID Connect extension', which is not relevant to the configuration needed for enabling user accounts. Lastly, the fifth node is about 'LDAP and Active Directory integration', which does not relate to OpenID Connect or enabling user accounts automatically in XWiki.

ContextualRecall: The score is 0.20 because while the 2nd node in the retrieval context mentions 'OpenId Connect Authentication' and 'xwiki.properties', which supports the context of adjusting the file for OpenID Connect in XWiki (sentence 1), the retrieval context lacks details on adding or updating specific properties (sentence 2), uncommenting and setting 'oidc.enableUser' to 'true' (sentence 3), saving the file and restarting XWiki (sentence 4), and the automatic activation of new user accounts (sentence 5).

CustomContextualRelevancy: The retrieval context does not mention updating the xwiki.properties file or the specific property oidc.enableUser, which are crucial details in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.45714285714285713

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.79 because the response addresses the main question about mapping additional user profile fields from the OpenID Connect provider to XWiki user properties. However, it includes examples related to custom directories and Google API, which are not directly relevant to the specific question, 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 discusses 'general OpenID Connect settings and endpoints' without mentioning 'user profile field mapping or customization'. The second node is about 'different authentication methods' and lacks information on 'mapping user profile fields from OpenID Connect'. The third node provides 'general information about the OpenID Connect Authenticator' but does not address 'user profile field mapping or customization'. The fourth node lists 'dependencies for the OpenID Connect extension' without relating to 'user profile field mapping or customization'. Lastly, the fifth node provides 'links to configuration guides and general information about the OpenID Connect Authenticator', but does not address 'user profile field mapping or customization'.

ContextualRecall: The score is 0.50 because while the 1st node in retrieval context supports the customization of user profile fields and mentions 'xwiki.properties' and 'oidc.user.mapping' (related to sentences 1 and 2), there is no specific information in the retrieval context about 'myxproperty1', 'myxproperty2', variable replacement, or the need to save and restart XWiki (related to sentences 3 and 4).

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5402173913043479

Individual_scores:

AnswerRelevancy: 0.15942028985507245

Faithfulness: 0.875

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.8333333333333334

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.16 because the majority of the output contains irrelevant statements that do not address the input about configuring group synchronization between XWiki and an OpenID Connect provider. The presence of numerous unrelated topics significantly lowers the relevancy score.

Faithfulness: The score is 0.88 because the actual output incorrectly states that the 'OpenID Connect' extension can be downloaded from the Wikia extension repository or via FTP/SFTP, which is not supported by the retrieval context.

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context, which discusses 'group synchronization, including the use of the 'oidc.groups.mapping' property', is ranked third. The first and second nodes focus on 'authentication management and OpenID Connect provider configuration' and 'container authentication', which do not mention group synchronization, and should be ranked lower than the relevant node.

ContextualRecall: The score is 0.83 because the retrieval context nodes provide relevant information for most of the expected output sentences, such as configuring group synchronization and mapping groups (nodes 1 and 3). However, it lacks details about saving the file and restarting the XWiki instance, which affects the completeness of the context.

CustomContextualRelevancy: The retrieval context mentions group synchronization and mapping using 'oidc.groups.mapping', but lacks details on 'oidc.groups.claim' and 'oidc.groups.separator'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.2

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because the output contains numerous irrelevant statements, including links to unrelated documentation and unclear sentences, none of which address the question about restricting file uploads on a wiki.

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 is about 'various administrative and application features in XWiki', none of which pertain to configuring file upload settings. The third node talks about 'PDF export customization using XSL and CSS', which does not relate to setting file upload restrictions. The fourth node lists 'various features and macros in XWiki', but does not address file upload configurations. The fifth and sixth nodes contain 'image links and references', which are unrelated to configuring file upload 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, as all instructions and configurations mentioned are missing.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_029_result.json

Overall_score: 0.546896551724138

Individual_scores:

AnswerRelevancy: 0.034482758620689655

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.03 because the actual output contains numerous irrelevant statements, such as links and information about XWiki documentation, PowerShell commands, and documentary reviews, none of which address the question about refusing script rights to a space administrator. The score is not zero because there may be a minimal connection or attempt to address the topic, albeit ineffectively.

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 is relevant as it discusses 'programming and script rights' and mentions that 'programming rights can only be granted from the wiki preferences page,' is ranked after an irrelevant node. The first node is about 'language preferences and translations in XWiki,' which is unrelated to script rights for administrators and should be ranked lower. Additionally, subsequent irrelevant nodes, such as the third node listing

various rights without addressing the specific question, are ranked too closely to the relevant node.
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 mentions that admin rights imply other rights and are not deniable, but it does not explicitly state that admin rights cannot be refused as in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.2

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because none of the statements in the actual output are relevant to configuring team spaces in XWiki as per the input request. The output contains numerous unrelated and nonsensical statements that do not address 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. The first node discusses a page in XWiki without mentioning access rights for teams, the second node talks about modifying resources using XML and HTTP methods, which is unrelated to team space permissions, and subsequent nodes continue to focus on unrelated topics such as editing comments, explaining spaces, actions after running a migrator, and configuration display options, none of which address configuring access rights for team spaces in XWiki.

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, indicating a complete lack of relevant information.

CustomContextualRelevancy: The retrieval context lacks information about creating groups, setting rights at wiki and space levels, and inheriting rights for child pages, which are crucial elements in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_031_result.json

Overall_score: 0.7495238095238095

Individual_scores:

AnswerRelevancy: 0.7333333333333333

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.7142857142857143

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.73 because the main content addresses the question about authenticating users with access tokens, but the inclusion of multiple source links that do not directly

contribute to the explanation reduces 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 0.50 because relevant nodes in the retrieval context are mixed with irrelevant ones. The first node discusses 'LDAP authentication settings in XWiki', which is unrelated to the input about token-based authentication, yet it is ranked higher than relevant nodes. Similarly, the third node talks about 'OpenID Connect scopes and response types', which are not directly related to JWT token-based authentication, but it is ranked above a relevant node. The relevant nodes that discuss 'Authorized Applications' and 'authentication using JWT tokens' are not consistently ranked higher than these irrelevant nodes.

ContextualRecall: The score is 0.71 because several key aspects of the expected output, such as authenticating users with JWTs and configuring the token authenticator, are well-supported by nodes in the retrieval context. However, details about generating a JWT token with specific claims and including it in the 'Authorization' header are not explicitly covered in the retrieval context.

CustomContextualRelevancy: The retrieval context includes details about configuring authorized applications and enabling the token authenticator, aligning with the expected answer. However, it lacks specific mention of required JWT claims and the use of the 'Authorization' header.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.38181818181818183

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.91 because the response effectively addresses the question about authenticating users with access tokens, but it includes 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 0.00 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 'details about the admin UI for model validation'. None of these are related to 'authenticating users with access tokens', leading to 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, as the 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 the 'Token-Authentifizierer'.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication, JWT, or the LLM-Anwendung (BETA)-Erweiterung mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.39069767441860465

Individual_scores:

AnswerRelevancy: 0.9534883720930233

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.95 because the answer is highly relevant and addresses the question about authenticating users with access tokens effectively. However, it includes source links that are not directly related to the topic, 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. The first node discusses 'upgrading XWiki' without mentioning user authentication with access tokens. The second node talks about 'features related to users and rights' but does not specifically address authentication with access tokens. The third node, while mentioning an 'Authentication guide', lacks information on authenticating users with access tokens or JWTs. Therefore, there are no relevant nodes ranked higher than irrelevant ones.

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 token-based authentication, JWTs, or related configuration and usage details.

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication or JWT, as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_qwen2_7b_Q4

File: qa_001_result.json

Overall_score: 0.37272727272727274

Individual_scores:

AnswerRelevancy: 0.8636363636363636

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.86 because the output mostly addresses the issue of the missing 'Bell' in XWiki, but includes irrelevant source links and dependency locations that do not directly 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 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 relate to 'enabling notifications in XWiki'. The second node describes 'the functionality and appearance of the notifications menu', but it does not address 'enabling notifications or fixing the bell icon issue'. The third node lists 'dependencies related to the notifications extension', which also fails to provide information on 'enabling notifications or fixing the bell icon issue'. Finally, the fourth node

explains 'how to override notification templates', which is unrelated to the issue at hand.

ContextualRecall: The score is 0.00 because the expected output about enabling notifications and setting `notifications.enabled` to `true` in `xwiki.properties` is not supported by any node(s) in the retrieval context.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.7611111111111111

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.8055555555555555

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

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 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 0.81 because most relevant nodes in the retrieval context are ranked higher, such as the first node explaining 'how to follow a user to receive notifications' and the third node mentioning 'settings menu for notifications'. However, the score is not higher because the second node, discussing 'filters and hidden pages', is irrelevant and should be ranked lower than the relevant nodes.

ContextualRecall: The score is 0.50 because some key details, like enabling notifications via `xwiki.properties` and subscribing to pages or users, are not found in the nodes in retrieval context, while other aspects like using the 'alert' menu and following users are partially supported by nodes 1 and 4.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 1.0

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

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 all relevant nodes are ranked higher than irrelevant ones. 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 is directly related to the input. The second node also aligns with the input 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.' Irrelevant nodes, such as the third one about 'administrators configuring notifications,' are ranked lower, maintaining perfect precision. Great job on the ranking!

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 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 user's notification settings.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

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. 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 is the fifth node, mentions the ability for users to change their own notification settings and the default auto-watched mode, but it is ranked lower than four irrelevant nodes. The first node discusses getting a notification RSS feed and general notification settings without addressing disabling notifications. The second node talks about following a user and default watch settings, lacking information on disabling notifications. The third node discusses filtering notifications and hidden pages without mentioning disabling notifications. The fourth node is about enabling and disabling email notifications at the server level, not directly related to user-level settings. These irrelevant nodes should be ranked lower than the relevant node.

ContextualRecall: The score is 0.50 because while nodes in the retrieval context support the idea of automatic page watching and accessing notification settings (related to sentences 1 and 3), they do not cover notifications for changes on one's work or deleting filters (related to sentences 2 and 4).

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.8509090909090908

Individual_scores:

AnswerRelevancy: 0.9545454545454546

Faithfulness: 0.9

ContextualPrecision: 1.0

ContextualRecall: 0.4

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant and provides useful information about configuring OpenID Connect in XWiki. However, it includes a mention of LDAP authentication, which is not relevant to the question asked.

Faithfulness: The score is 0.90 because the actual output incorrectly claims that user creation on login is enabled by setting a property in the xwiki.security.authorisation configuration file, whereas the retrieval context specifies that it should be set in xwiki.cfg.

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. All irrelevant nodes discussing topics like 'domain-based authentication and session cookies', 'installation instructions', 'container authentication and LDAP', and 'LDAP configuration' are ranked lower, ensuring precision in retrieval.

ContextualRecall: The score is 0.40 because while the 1st node in the retrieval context supports the key point about enabling user accounts on first login, it lacks information on updating the 'xwiki.properties' file, saving it, restarting XWiki, and ensuring automatic account enablement via OpenID Connect.

CustomContextualRelevancy: The retrieval context includes the exact information about the 'oidc.enableUser' property being set to true, which is the key detail in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.9647619047619047

Individual_scores:

AnswerRelevancy: 0.9666666666666667

Faithfulness: 0.8571428571428571

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.97 because the response is highly relevant and directly addresses the question about customizing user profile fields from an OpenID Connect provider in XWiki. The only minor issue is the inclusion of a source link, which is not directly necessary for the input question.

Faithfulness: The score is 0.86 because the actual output incorrectly claims that XWiki uses an AuthenticationHandler class, while the retrieval context specifies that the authentication management class is org.xwiki.contrib.oidc.auth.OIDCAuthServiceImpl.

ContextualPrecision: The score is 1.00 because the relevant node in the retrieval context is ranked first and directly addresses the input query by stating 'It's possible to associate non standard properties coming from the OpenID Connect provider with the XWiki user' and providing an example of 'oidc.user.mapping=myxproperty1=\${oidc.user.subject}', which is exactly about customizing user profile fields. The irrelevant nodes, such as the second node discussing access to various claims like 'profile', 'email', 'address', and 'phone', and the third node focusing on OpenID Connect provider endpoints and HTTP headers, are correctly ranked lower as they do not pertain to the customization of user profile fields.

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

CustomContextualRelevancy: The retrieval context includes detailed instructions on mapping user profile fields using the `oidc.user.mapping` property, matching the expected output's explanation of customizing user profile fields in the `xwiki.properties` file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.8856410256410256

Individual_scores:

AnswerRelevancy: 0.9666666666666667

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.46153846153846156

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.97 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 that doesn't directly address the process, 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 all relevant nodes in the retrieval context are ranked higher than the irrelevant nodes. The first three nodes provide detailed information on configuring group synchronization, specifying the claim for group lists, and mapping XWiki groups to provider groups, which are directly relevant to the input. The irrelevant nodes, ranked fourth and fifth, discuss general OpenID Connect authentication and provider endpoints, which are not directly related to the group synchronization process.

ContextualRecall: The score is 0.46 because while nodes in the retrieval context provide detailed information about configuring properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, they lack explicit instructions for actions such as locating and editing the `xwiki.properties` file, replacing placeholders with actual values, and restarting the XWiki instance, which are crucial steps 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 the `xwiki.properties` file and restarting XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6318954248366013

Individual_scores:

AnswerRelevancy: 0.9705882352941176

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.5555555555555556

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.97 because the response is highly relevant and directly addresses the question about setting upload restrictions, with only a minor irrelevant mention of a general guide link.

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 about setting the maximum size of an attachment in XWikiPreferences, which is directly related to configuring the maximum upload size', is ranked third. However, irrelevant nodes such as the first node, which 'discusses linking to external files and images, but does not address configuring upload restrictions or size limits', and the second node, which 'provides a general overview of XWiki's structure and features, but does not mention configuring attachment upload restrictions or size limits', are ranked higher, reducing the score.

ContextualRecall: The score is 0.56 because while the 3rd node in the retrieval context supports the steps for configuring the maximum attachment size as described in sentences 5-9 of the expected output, there is no mention of configuring allowed mimetypes or ensuring only images or PDF files can be uploaded, which are key elements of sentences 1-4.

CustomContextualRelevancy: The retrieval context mentions setting maximum attachment size and mimetype restrictions but lacks specific instructions on configuring mimetypes in xwiki.properties and setting size to 10MB.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.7853658536585366

Individual_scores:

AnswerRelevancy: 0.926829268292683

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is highly relevant and effectively addresses the question about denying script rights to a space administrator. However, it includes a repeated irrelevant statement about a source URL, which 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. 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 default states 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 address the specific question about denying script rights to a space administrator. The fourth node provides technical details about checking access rights but does not relate to denying script rights to a space administrator. The fifth node discusses converting rights during page migration but does not address 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.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.6725641025641025

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 0.8461538461538461

ContextualPrecision: 0.6666666666666666

ContextualRecall: 0.75

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.80 because the output provides relevant information about configuring permissions in XWiki, but it includes multiple source links and examples that do not directly address the specific configuration question, reducing the overall relevancy.

Faithfulness: The score is 0.85 because the actual output inaccurately claims that XWiki uses roles to manage permissions and mentions a Public Role for default access, which are not specified in the retrieval context.

ContextualPrecision: The score is 0.67 because relevant nodes in retrieval context, such as the first node discussing 'rights inheritance and administration features' and the fourth node explaining 'XWiki's permission system', are ranked higher. However, irrelevant nodes like the second node focusing on 'parent/child relationship and Nested Pages', and the third node about 'string formatting and OpenID Connect properties', should be ranked lower to improve the score.

ContextualRecall: The score is 0.75 because while the retrieval context nodes provide a strong foundation for understanding the general process of setting rights (nodes 4 and 6), specific details about all teams viewing all spaces and each team editing their own space are not directly covered.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.5133333333333333

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.90 because the response is highly relevant to the input, addressing the issue of the unavailable bell icon. However, it includes a source link that does not directly contribute to solving the problem, 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 0.17 because the relevant node in the retrieval context, which is the sixth node, mentions 'editing the xwiki.properties file' to enable notifications, is ranked lower than five irrelevant nodes. These irrelevant nodes, such as the first node that 'discusses custom notification

filters and their behavior but does not address enabling notifications in the wiki settings', should be ranked lower. Similarly, the second node 'describes the Alert menu and the display of notifications but does not mention enabling notifications through settings', and the third node 'lists different ways to receive notifications but does not provide information on enabling notifications in the wiki settings', should also be ranked lower. This misordering results in a low precision score.

ContextualRecall: The score is 0.50 because while the 5th node in the retrieval context mentions 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 the `xwiki.properties` file or setting `notifications.enabled` to `true`.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3888888888888889

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant to the issue of not receiving notifications, with only a minor irrelevant statement about a source link that does not detract significantly from the overall helpfulness.

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 to various projects, which do not relate to enabling notifications or settings in xwiki.properties.' The second node also consists of 'image links and references to projects, which are irrelevant to the topic of notifications or user settings.' Similarly, the third node includes 'image links and references that do not pertain to the process of enabling notifications or following users in a wiki.'

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 only contain image links and references.

CustomContextualRelevancy: The retrieval context contains only image links and does not provide any information related to enabling notifications or following users/pages as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.3555555555555555

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.78 because the answer 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 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, 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 'importing documents, sharing pages, and other functionalities but not about notifications or user activity settings.' This results in irrelevant nodes being ranked higher, 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 key elements in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.36

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.80 because the response addresses the main question about disabling notifications, but it includes several irrelevant URLs related to XWiki documentation, syntax, extensions, security module, and features, which do not pertain to the specific 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 'attaching files', which is unrelated to 'notification settings'. The second node is about 'moving and deleting attachments', not 'managing notification settings'. The third node covers 'displaying pages and macros', which does not pertain to 'notification settings'. The fourth node is about 'administrating wiki features', with no mention of 'notification settings'. The fifth node discusses 'permissions and security', not 'notification settings'. Lastly, the sixth node lists 'applications and features 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 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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.5700000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.6

CustomContextualRelevancy: 0.0

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 0.25 because the relevant node in the retrieval context, which discusses the `xwiki.properties` file necessary for enabling automatic user activation with OpenID Connect, is ranked fourth. The first three nodes are irrelevant as they discuss 'domain-based authentication and session cookie configuration', 'general information about OpenID Connect without mentioning specific properties', and 'metadata about the OpenID Connect Authenticator without addressing automatic user activation'. These irrelevant nodes should be ranked lower than the relevant node.

ContextualRecall: The score is 0.60 because while the 4th node in the retrieval context supports the instructions to modify 'xwiki.properties' and set 'oidc.enableUser' to 'true', there is no explicit mention in any node about restarting the XWiki instance or ensuring new user accounts are automatically activated upon first login.

CustomContextualRelevancy: The retrieval context does not contain any information about updating the xwiki.properties file or the property oidc.enableUser, which is crucial for the expected answer.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.6618181818181819

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.91 because the response is highly relevant to the input question, providing useful information on customizing user profile fields in XWiki. However, it includes source links that do not directly address the question, 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 1.00 because the relevant node in the retrieval context, which mentions the 'xwiki.properties' file for configuring OpenID Connect settings in XWiki, is ranked first. This ensures that the most pertinent information is prioritized over irrelevant nodes discussing topics like container authentication and LDAP, general OpenID Connect guides, the authenticator extension's compatibility, and its dependencies, which do not address user profile field mapping.

ContextualRecall: The score is 0.20 because only the 1st sentence in the expected output is supported by the 1st node in the retrieval context, which mentions 'xwiki.properties' and OpenID Connect. The rest

of the sentences lack corresponding information in the retrieval context, such as details on 'oidc.user.mapping', examples of custom property mapping, or instructions on saving and restarting XWiki.

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.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.4

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

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. 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 discusses 'generic OpenID Connect provider configurations and endpoints' but lacks specific information on group synchronization setup. The second node is about 'different authentication methods' and does not mention group synchronization. The third node provides links to configuration guides but does not address the specific steps 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 details on 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 does not contain any information about configuring group synchronization in xwiki.properties, claims, separators, or mappings as detailed in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6814285714285714

Individual_scores:

AnswerRelevancy: 0.85

Faithfulness: 0.8571428571428571

ContextualPrecision: 0.325

ContextualRecall: 0.875

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.85 because the output mostly addresses the question about configuring attachment settings, but includes irrelevant links to pages about exports, XWiki syntax, and document lifecycle, which do not directly relate to the query.

Faithfulness: The score is 0.86 because the actual output incorrectly states that the maximum upload size is 10 MB, while the retrieval context specifies a default maximum size of 100GB.

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 and PDF export features', and 'various macros and features', none of which relate to configuring attachment upload restrictions in a wiki. The relevant nodes, which discuss 'Size Limit' and 'Mimetype Restriction', and provide 'detailed steps on setting the maximum size of an attachment', are ranked fourth and fifth. This lowers the precision score as the relevant information is not prioritized higher than the irrelevant nodes.

ContextualRecall: The score is 0.88 because while the 5th node in the retrieval context provides detailed steps for changing the maximum size of an attachment, similar to sentences in the expected output, it does not mention configuring MIME types, which is a key part of the expected output.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.3764705882352941

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the response is mostly relevant to the question about denying script rights to a domain administrator, but it includes source links that do not directly address 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 0.00 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 'templates and page creation', and the third node covers 'custom configuration and properties'. None of these topics relate to denying script rights to an administrator, which is the focus of the input.

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

CustomContextualRelevancy: The retrieval context does not contain any information about admin rights or their implications in XWiki, which is essential to the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.385

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.125

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.80 because the response addresses the main question about configuring team access in XWiki effectively, but it includes several irrelevant source links that do not directly relate to the specific task of configuring team access.

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 compatibility', which is unrelated to configuring access control for team areas. The second node is about 'notifications and application details', which does not pertain to setting up access controls in XWiki. The third node explains 'nested pages and subwikis', which is not directly relevant to configuring team-specific access rights. The fourth node provides information about 'managing users and groups', but it does not specifically address the configuration of access rights for team areas. The fifth node discusses 'group synchronization and customization', which does not directly relate to setting access controls for team areas in XWiki.

ContextualRecall: The score is 0.12 because only sentence 1 in the expected output is partially supported by node 4 in the retrieval context, which mentions managing users and groups. The rest of the sentences in the expected output are not supported by any nodes in the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.4818181818181818

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.91 because the response is mostly relevant and addresses the issue of the notification bell not appearing. However, it includes source links and documentation references that do not directly solve the problem, 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 'custom notification filters and their behavior', the second node lists 'various links related to notification settings and customization', the third node provides 'links to various notification preferences', the fourth node describes 'the notification application and its features', and the fifth node talks about 'getting a notification RSS feed and notification settings'. None of these nodes address the issue of enabling notifications in the wiki settings, which is crucial for the input query.

ContextualRecall: The score is 0.50 because while the 1st sentence in the expected output is supported by the 1st node in the retrieval context, the 2nd sentence lacks direct support as none of the nodes mention setting 'notifications.enabled' in 'xwiki.properties'.

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 answer.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.7492307692307693

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.4

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 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 ranked higher than irrelevant nodes. The first node provides information on ensuring notifications are activated, and the second node aligns with following users for notifications. The third node, discussing RSS feeds and notification settings, and the fourth node, listing various notification preferences, do not directly address enabling notifications or following users and are correctly ranked lower.

ContextualRecall: The score is 0.50 because while nodes in the retrieval context support some key concepts like enabling notifications and following users (related to sentences 1, 2, 3, and 4), they lack specific details such as setting 'notifications.enabled' in 'xwiki.properties', the alert menu, receiving all notifications from a user, and the user profile network tab (related to sentences 2, 3, 5, and 6).

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.8436363636363637

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.82 because the response mostly addresses the question about enabling notifications for one's own activity, but includes irrelevant source links that do not directly relate 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, '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 the irrelevant nodes, discussing topics like RSS feeds, general notification settings, and technical aspects, are ranked lower.

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 information about disabling the filter in 'Options de filtrage avancées' is not found in 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 option to change this setting. However, it does not explicitly mention the 'Options de filtrage avancées' section.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.86

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.75 because the response addresses the question about disabling notifications, but includes irrelevant links that do not directly help with the issue.

Faithfulness: The score is 1.00 because there are no contradictions, indicating perfect alignment 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 irrelevant node, ranked fourth, discusses 'RSS feeds and notification types', which do not directly address the issue of disabling automatic notifications and modifying watch settings. Great job on keeping the relevant information at the top!

ContextualRecall: The score is 0.75 because the expected output sentences about default auto-watched mode and modifying settings are supported by nodes in the retrieval context, specifically nodes 1 and 2. However, the sentence regarding notifications about changes lacks direct support from the retrieval context.

CustomContextualRelevancy: The retrieval context includes information on 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 notifications for changes on one's work and modifying settings in the user profile.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.47411764705882353

Individual_scores:

AnswerRelevancy: 0.9705882352941176

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.97 because the response is highly relevant to configuring OpenID Connect for automatic user activation in XWiki, with only a minor irrelevant mention of LDAP Authenticator.

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 instances and session cookie configuration', which does not pertain to enabling user accounts automatically with OpenID Connect. The second node focuses on 'container authentication and LDAP', lacking the necessary details for automatic user activation with OpenID Connect. The third node, although mentioning OpenID Connect, only covers 'installation and general configuration guides', missing the specific property for automatic activation. The fourth node lists 'dependencies for the OpenID Connect extension', unrelated to the configuration for user activation. Lastly, the fifth node is about 'LDAP and Active Directory integration', not relevant to the OpenID Connect configuration needed.

ContextualRecall: The score is 0.40 because while there are relevant mentions of OpenID Connect and the xwiki.properties file in nodes 1 and 3 in the retrieval context, the specific instructions for adding or updating the property for user account activation, saving the file, and restarting XWiki are not covered.

CustomContextualRelevancy: The retrieval context lacks specific information about adjusting xwiki.properties to activate user accounts with OpenID Connect, as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.48

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question about mapping additional user profile fields from an OpenID Connect provider to XWiki user properties. 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 does not mention user profile field mapping or customization. The second node provides information about authentication methods and extensions, but does not address customizing user profile fields or mapping them. The third node is about the OpenID Connect Authenticator extension, its developer, and installation details, but does not mention user profile field mapping. The fourth node lists dependencies for the OpenID Connect Authenticator extension and does not discuss user profile field mapping. The fifth node provides configuration guides for specific systems and general information about the OpenID Connect Authenticator, but does not mention mapping user profile fields.

ContextualRecall: The score is 0.40 because only some parts of the expected output, such as saving 'xwiki.properties' and restarting XWiki, are supported by the nodes in retrieval context, while key details

about 'oidc.user.mapping' and specific examples are not directly supported by any node.
CustomContextualRelevancy: The retrieval context lacks information about customizing user profile fields using 'oidc.user.mapping' in 'xwiki.properties'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.5837037037037037

Individual_scores:

AnswerRelevancy: 0.8518518518518519

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.85 because the response addresses the main question about configuring group synchronization between XWiki and OpenID Connect, but includes some irrelevant details about accessing the XWiki configuration and administration interface, which are not directly related to 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 0.33 because the third node in the retrieval context is relevant as it 'directly addresses group synchronization by mentioning the use of the 'oidc.groups.mapping' property and the need to configure it in the 'xwiki.properties' file.' However, the first, second, fourth, and fifth nodes are ranked higher or similarly to the relevant node, despite being irrelevant. The first node discusses 'general OpenID Connect properties and endpoints' without mentioning group synchronization, the second node focuses on 'authentication classes' rather than group synchronization, the fourth node provides 'general information about OpenID Connect and its limitations' without covering group synchronization steps, and the fifth node lists 'dependencies for the OpenID Connect extension' unrelated to group synchronization configuration.

ContextualRecall: The score is 0.33 because while the 3rd node in the retrieval context covers the general aspects of configuring group synchronization and mentions relevant 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, providing examples, and instructions on saving and restarting XWiki.

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. It also does not cover editing 'xwiki.properties' or restarting XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3583333333333333

Individual_scores:

AnswerRelevancy: 0.7916666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.79 because the response partially addresses the question about setting file upload restrictions on a wiki, 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 'language preferences and translations in XWiki', which is unrelated to configuring file upload restrictions. The second node is about 'various applications and features in XWiki, such as administration and mail configuration', not about file upload settings. The third node covers 'PDF export customization using CSS and XSL', which is not relevant to setting upload restrictions for images and PDFs. The fourth node lists 'various macros and features related to notifications, office document viewing, and PDF viewing', which do not pertain to file upload restrictions. The fifth node contains 'links to various references and images', which do not provide information on configuring file upload restrictions. Finally, the sixth node includes 'references to various entities and their logos', which are not related 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 answer.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.7480952380952381

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.86 because the response mostly addresses the question about refusing script rights to a space administrator, but it includes unrelated source links that slightly detract from its relevance.

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

ContextualPrecision: The score is 0.58 because the first node in the retrieval context is about 'language preferences and translations in XWiki,' which is unrelated to the question about script rights, yet it is ranked higher than relevant nodes. Additionally, the fourth node discusses 'checking access rights programmatically,' and the fifth node talks about 'executing code with different user rights,' both of which do not directly address the question and are ranked higher than they should be. However, the second and third nodes are relevant as they discuss 'programming and script rights in XWiki' and provide a 'detailed list of rights,' respectively, which are pertinent to the input question.

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 cannot be denied. 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.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_030_result.json

Overall_score: 0.3733333333333335
Individual_scores:
AnswerRelevancy: 0.8666666666666667
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.87 because the response effectively addresses the question about configuring access and modification settings in XWiki, 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 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 is about 'sending representations and modifying resources,' which is not relevant to setting access permissions for team spaces in XWiki. The third node addresses 'editing comments and attachments,' which does not address the question of configuring team 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 discusses 'actions needed after running a migrator,' which does not relate to configuring team access permissions in XWiki. Lastly, the sixth node describes 'configuration display settings in XWiki,' which is not relevant to the question about team access rights.

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 creating groups, defining rights, or inheriting rights as outlined in the expected output.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_031_result.json

Overall_score: 0.775270935960591
Individual_scores:
AnswerRelevancy: 0.8620689655172413
Faithfulness: 1.0
ContextualPrecision: 0.5
ContextualRecall: 0.7142857142857143
CustomContextualRelevancy: 0.8
Reasons:

AnswerRelevancy: The score is 0.86 because the response mostly addresses the question about authenticating users with access tokens, but it includes multiple 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 'provides information about configuring authorized applications, which is relevant to using access tokens for authentication,' and the fourth node 'describes the extension that allows authentication using JWT,' both supporting the input query. However, the first node discusses 'LDAP configuration and local account login settings, which are unrelated to token-based authentication,' and the third node 'lists claims that can be requested in access tokens, but it does not directly relate to the process of authenticating users with access tokens,' which should be ranked lower. Additionally, the fifth node 'discusses general authentication mechanisms in XWiki, which does not specifically address the use of access tokens or JWT for authentication,' further affecting the precision score.

ContextualRecall: The score is 0.71 because while several key aspects such as authenticating users with JWTs, enabling the token authenticator, and configuring authorized applications are well-supported by nodes in the retrieval context, details about generating a JWT token with specific claims and including it in the 'Authorization' header are missing.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_result.json

Overall_score: 0.3555555555555555

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.78 because the response mostly addresses the question about authenticating users with access tokens, but includes source links that do not directly relate 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 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 'token-based authentication'. The third node provides information about the 'admin UI for model validation', which does not relate 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, JWT, or the LLM-Anwendung (BETA) extension mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.32

Individual_scores:

AnswerRelevancy: 0.6

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.60 because the response partially addresses the question about authenticating users with access tokens, but includes irrelevant source links and information on access rights management 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 'upgrading XWiki' without mentioning 'user authentication or access tokens'. The second node focuses on 'features of the Administration Application', specifically 'users, groups, and rights', but lacks any reference to 'authentication with access tokens'. The third node is about 'rights management and user profiles', which also does not address '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, indicating a complete lack of relevant information.

CustomContextualRelevancy: The retrieval context lacks any information related to token-based authentication or JWT, which are central 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: 0.6

Reason: The score is 0.60 because the summary fails to address some questions that the original text can answer, indicating it lacks completeness and omits certain details.

Score_breakdown:

Alignment: 1.0

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.8181818181818182

Reason: The score is 0.82 because the summary contains a contradiction regarding the scope of the Programming Right and includes extra information about a tabular view that is not present in the original text. Despite these issues, the summary is mostly accurate and informative.

Score_breakdown:

Alignment: 0.8181818181818182

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.5555555555555556

Reason: The score is 0.56 because the summary includes extra information not found in the original text, such as details about wiki access implications, sub-wiki access rights, terminal page permissions, and additional resources for understanding permissions. Additionally, the summary fails to answer a question that the original text can, indicating a lack of completeness and accuracy.

Score_breakdown:

Alignment: 0.5555555555555556

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.5555555555555556

Reason: The score is 0.56 because the summary includes extra information not present in the original text, such as troubleshooting XWiki, Debian packages, demo packages, downgrading, and exporting content as XAR files. Additionally, the summary fails to answer questions about XWiki's migration paths and automatic database schema upgrades, which the original text can address.

Score_breakdown:

Alignment: 0.5555555555555556

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.8

Reason: The score is 0.80 because the summary accurately reflects the original text without contradictions or unnecessary additions. However, it lacks completeness as it does not address all questions that the original text can answer, such as the availability of a feed for all Blog posts.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains contradictions regarding developer resources and includes extra information about user actions not specified in the original text. Additionally, it fails to answer questions that the original text can address, such as guidance on creating pages and coverage of basic use cases.

Score_breakdown:

Alignment: 0.6

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.875

Reason: The score is 0.88 because the summary includes extra information about a 'Version Control page' in XWiki that is not mentioned in the original text. Overall, the summary is still quite accurate and closely aligned with the original text.

Score_breakdown:

Alignment: 0.875

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, such as images illustrating user rights settings and links for visual guidance. Additionally, the summary fails to address a question about configurations in XWiki that the original text can answer.

Score_breakdown:

Alignment: 0.6

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 not mentioned in the original text, such as references to XWiki and its pages. 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: 0.8333333333333334

Reason: The score is 0.83 because the summary includes extra information by providing specific examples of applications like blogs or task managers, which were not mentioned in the original text. However, there are no contradictions, indicating a generally accurate representation of the original content.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.GPT4o-mini

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 users typically starting with a single wiki and setting different panels for various spaces within a wiki. Additionally, the summary fails to answer a question that the original text can, specifically about exporting a page in PDF format in XWiki.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary contains a significant contradiction regarding the scope of the Programming Right, 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 moderate level of accuracy and completeness in the summary.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.7692307692307693

Reason: The score is 0.77 because the summary includes some contradictions regarding terminal pages and access rights, and it adds extra information about email verification and additional resources that are not in the original text. Despite these issues, the summary is relatively accurate and informative.

Score_breakdown:

Alignment: 0.7692307692307693

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.7333333333333333

Reason: The score is 0.73 because the summary includes extra information not present in the original text, such as specific methods and commands for upgrading, and it fails to answer a question that the original text can address. However, there are no contradictions, indicating a relatively good alignment with the original text.

Score_breakdown:

Alignment: 0.7333333333333333

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.8

Reason: The score is 0.80 because the summary accurately reflects the original text without contradictions or unnecessary extra information. However, it misses answering a specific question that the original text addresses, indicating room for improvement in comprehensiveness.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains contradictions, such as incorrectly stating that the Extensions wiki is a resource for developers, and includes extra information not mentioned in the original text, like specific actions requiring login and a 'Creating a page' guide. Additionally, the summary fails to answer questions that the original text can, such as whether developers can build complex web applications using XWiki.

Score_breakdown:

Alignment: 0.4

Coverage: 0.8

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 omits details that the original text can answer, such as the presence of a 'History' tab in XWiki.

Score_breakdown:

Alignment: 0.8888888888888888

Coverage: 0.8

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 found 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.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.6

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 extraneous information.

Score_breakdown:
Alignment: 1.0
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 includes extra information about blogs and task managers that is not specifically 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

Model: AI.Models.claude3_5_sonnet

File: summ_001_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 some details, such as the ability to export a page in PDF format in XWiki, which slightly affects its comprehensiveness.
Score_breakdown:
Alignment: 1.0
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 contradictions regarding where programming permissions can be granted and the order of permission checking, as well as extra information about a detailed table that is not present in the original text. Additionally, the summary fails to answer questions about the default status of the edit right and the inclusivity of the administration right, which the original text can answer.
Score_breakdown:
Alignment: 0.8125
Coverage: 0.6
Input_language: en
Summary_language: en

File: summ_003_result.json

Score: 0.38461538461538464
Reason: The score is 0.38 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: 0.38461538461538464
Coverage: 1.0
Input_language: en
Summary_language: en

File: summ_004_result.json

Score: 0.8571428571428571
Reason: The score is 0.86 because the summary includes extra information not present in the original text, such as details about the database schema being automatically upgraded and a preference for upgrading over downgrading. However, there are no contradictions, indicating a generally accurate representation of the original text.
Score_breakdown:
Alignment: 0.8571428571428571
Coverage: 1.0
Input_language: en
Summary_language: en

File: summ_005_result.json

Score: 1.0
Reason: The score is 1.00 because the summary perfectly matches the original text without any contradictions or unnecessary additions. Excellent summarization quality.
Score_breakdown:
Alignment: 1.0
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 extra information not present in the original text, such as regular users being able to create and edit pages and access page history, and it lacks details about guides for users on creating a page and additional applications on the Extensions wiki.
Score_breakdown:
Alignment: 0.6666666666666666
Coverage: 0.6
Input_language: en
Summary_language: en

File: summ_007_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. Excellent summarization quality.
Score_breakdown:
Alignment: 1.0
Coverage: 1.0
Input_language: en
Summary_language: en

File: summ_008_result.json

Score: 0.8333333333333334

Reason: The score is 0.83 because the summary contains some contradictions regarding account creation and viewing/editing permissions, which slightly impacts its accuracy. However, there is no extra information or unanswered questions, indicating a generally good summarization quality.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.8333333333333334

Reason: The score is 0.83 because the summary includes extra information such as 'knowledge bases', 'documentation', 'websites', 'intranets', or 'extranets' that were 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_010_result.json

Score: 0.6

Reason: The score is 0.60 because the summary includes contradictions, such as claiming XWiki can be used for educational purposes, which is not supported by the original text. Additionally, the summary introduces extra information about XWiki being used as a community hub, presentation website, or public wiki, which is not mentioned in the original text.

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: 0.2

Reason: The score is 0.20 because the summary contains significant contradictions with the original text regarding the basic units of content and the display of panels, includes extra information not present in the original text, and fails to answer questions that the original text can address. These issues indicate a lack of accuracy and completeness in the summary.

Score_breakdown:

Alignment: 0.2

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.2

Reason: The score is 0.20 because the summary contains multiple contradictions with the original text, such as incorrect information about permissions and rights. Additionally, it includes extra information not present in the original text and fails to answer several questions that the original text can address. These issues significantly reduce the accuracy and completeness of the summary.

Score_breakdown:

Alignment: 0.5263157894736842

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.14285714285714285

Reason: The score is 0.14 because the summary introduces several pieces of extra information not present in the original text, such as details about Main Wiki, Sub-Wiki, and Page levels, user/group definitions, and specific permission-setting processes. Additionally, the summary fails to address questions that the original text can answer, like permission configurations and public/private wiki options in XWiki.

Score_breakdown:

Alignment: 0.14285714285714285

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.0

Reason: The score is 0.00 because the summary includes extensive extra information not present in the original text, indicating a significant deviation from the source material.

Score_breakdown:

Alignment: 0.0

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_005_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, indicating a moderate level of accuracy and completeness.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.6

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 user capabilities like creating, editing, or viewing pages, and accessing page history. Additionally, it inaccurately states that the text provides links to documentation for users, administrators, and developers. Furthermore, the summary fails to address specific questions that the original text can answer, such as the availability of guides for creating a page, pre-installed applications for developers, and a dedicated wiki for XWiki Rendering documentation.

Score_breakdown:

Alignment: 0.5

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains contradictions regarding the functionality of the 'History' tab and the process of reverting to previous versions, as well as extra information about a 'Version Control page' not present in the original text. Additionally, the summary fails to address a question about content saving that the original text can answer.

Score_breakdown:

Alignment: 0.4

Coverage: 0.8

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, such as screenshots illustrating user rights configurations and a guide on creating a basic application in XWiki. Additionally, the summary fails to answer questions about XWiki's capabilities regarding user account creation in a locked registration configuration and admin options in a protected wiki configuration, which the original text can address.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.4

Reason: The score is 0.40 because the summary includes extra information not present in the original text and fails to answer several questions that the original text can address. This indicates a lack of coverage and accuracy in the summary.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_010_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 lacks some detail, as it cannot answer whether XWiki is used for a variety of use cases such as intranets and public websites.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

Model: AI.Models.gemma2_9B_Q4

File: summ_001_result.json

Score: 0.5

Reason: The score is 0.50 because the summary contains contradictions regarding the features of pages and editing options, includes extra information not present in the original text, and fails to answer questions about subwikis and exporting pages in PDF format that the original text can address.

Score_breakdown:

Alignment: 0.5

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.2

Reason: The score is 0.20 because the summary contains multiple contradictions with the original text, such as incorrect details about the 'Script' permission and the order of permission checks. Additionally, it includes extra information not present in the original text and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.625

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains a contradiction regarding the implicit denial of rights when explicitly set, includes extra information about specific permission types and user groups not present in the original text, and fails to answer questions about overriding permissions and setting access rights at a terminal page level.

Score_breakdown:

Alignment: 0.625

Coverage: 0.6

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 with the original text, such as misrepresenting the recommended methods for upgrading and downgrading XWiki. Additionally, it includes a substantial amount of extra information not present in the original text, such as details about manual upgrades and community resources. Furthermore, the summary fails to answer questions that the original text can address, indicating a lack of coverage and accuracy.

Score_breakdown:

Alignment: 0.2

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.9090909090909091

Reason: The score is 0.91 because the summary is mostly accurate with only minor discrepancies, such as including extra information about a Watchlist application requirement not mentioned in the original text.

Score_breakdown:

Alignment: 0.9090909090909091

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.2

Reason: The score is 0.20 because the summary includes extra information not present in the original text and fails to answer several questions that the original text can address, indicating a significant deviation from the original content.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 0.2

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, such as an 'XWiki Version Control page,' and fails to answer a question that the original text can address regarding whether all content added to XWiki is saved over time. Despite these issues, the summary is largely accurate, reflecting a good understanding of the original text.

Score_breakdown:

Alignment: 0.8888888888888888

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.4444444444444444

Reason: The score is 0.44 because the summary introduces additional information not present in the original text and contains contradictions regarding user account creation permissions. These issues indicate a lack of fidelity to the original content.

Score_breakdown:

Alignment: 0.4444444444444444

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 includes extra information about wikis being used for knowledge bases, documentation, and communication platforms that is not mentioned in the original text, and it fails to address whether most wikis include features like access rights management and notifications.

Score_breakdown:

Alignment: 0.8

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_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. Excellent summarization quality.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.llama3_1_402b

File: summ_001_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 the specification of creating wikis for separate types of content and setting panels differently for each space. However, there is no contradicting information, which is positive for the summary's quality.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.2

Reason: The score is 0.20 because the summary contains a significant contradiction regarding the priority order of rights, includes extra information not present in the original text, and fails to answer several questions that the original text can address. These issues indicate a low-quality summarization.

Score_breakdown:

Alignment: 0.7692307692307693

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.25

Reason: The score is 0.25 because the summary contains significant contradictions and includes a substantial amount of extra information not present in the original text, indicating a lack of accuracy and relevance.

Score_breakdown:

Alignment: 0.25

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.38461538461538464

Reason: The score is 0.38 because the summary contains multiple contradictions with the original text, such as incorrect advice on data backup and methods for upgrading and downgrading. It also includes extra information not present in the original text, like the claim of providing a comprehensive guide and specific instructions for admin login. Additionally, the summary fails to answer questions that the original text can, indicating a lack of coverage and accuracy.

Score_breakdown:

Alignment: 0.38461538461538464

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 contains a minor contradiction regarding the installation requirement for the Watchlist Feature, which slightly affects its accuracy. However, there is no extra information or unanswered questions, 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: 0.5714285714285714

Reason: The score is 0.57 because the summary includes extra information not present in the original text, such as the guide providing links to tutorials and the conclusion pointing to the next step. This

indicates a moderate level of accuracy and completeness in the summarization.

Score_breakdown:

Alignment: 0.5714285714285714

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 is mostly accurate and concise, with no contradictions or unnecessary information added. However, it lacks the ability to answer a specific question that the original text can address, indicating a slight gap in comprehensiveness.

Score_breakdown:

Alignment: 1.0

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 multiple contradictions regarding configurations and permissions that are not accurately reflected in the original text. Additionally, the summary includes extra information not present in the original text, such as screenshots and a comprehensive guide, which affects its accuracy and completeness.

Score_breakdown:

Alignment: 0.5

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.6

Reason: The score is 0.60 because the summary fails to address some questions that the original text can answer, indicating that it lacks completeness in capturing all relevant details.

Score_breakdown:

Alignment: 1.0

Coverage: 0.6

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 not present in the original text and fails to address a question that the original text can answer. 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

Model: AI.Models.llama3_1_8b_Q4

File: summ_001_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains multiple contradictions with the original text, such as the location of the tabbed area and the scope of actions possible. Additionally, it includes extra information not found in the original text, like XWiki being a collaborative content management system and the use of wikis for content separation. Furthermore, the summary fails to answer several questions that the original text can address, such as editing capabilities and export options in XWiki.

Score_breakdown:

Alignment: 0.4444444444444444

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.4666666666666667

Reason: The score is 0.47 because the summary contains contradictions regarding the delete permissions, includes extra details about a permissions table not present in the original text, and fails to answer questions about default statuses and specific rights that the original text can address.

Score_breakdown:

Alignment: 0.4666666666666667

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.25

Reason: The score is 0.25 because the summary contains significant contradictions and extra information not present in the original text. Additionally, there are several questions that the original text can answer but the summary cannot, indicating a lack of coverage and accuracy in the summary.

Score_breakdown:

Alignment: 0.25

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.16666666666666666

Reason: The score is 0.17 because the summary includes several pieces of extra information not present in the original text, such as details about the Distribution Wizard, exporting and importing content, and Solr configuration. Additionally, the summary fails to address specific questions that the original text can answer, such as recommendations for checking release notes and details about the upgrade process, which indicates a significant lack of coverage and accuracy.

Score_breakdown:

Alignment: 0.16666666666666666

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 contains a contradiction regarding the default RSS feeds and the Watchlist Feature, which slightly impacts the accuracy. However, there is no extra information added, indicating a generally high-quality summarization.

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 specific login requirements for actions and detailed resources for developers, which affects its accuracy and completeness.

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 is mostly accurate and concise, with no contradictions or unnecessary extra information. However, it misses answering a specific question that the original text can address.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.75

Reason: The score is 0.75 because the summary includes contradictions regarding specific configurations not mentioned in the original text and adds extra information about setting user rights not covered in the original. Additionally, it fails to address a question about admin-only configurations that the original text can answer.

Score_breakdown:

Alignment: 0.75

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 information. However, it lacks some detail, as it cannot answer specific questions that the original text can, such as whether most wikis include features like access rights management.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

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 specific applications like blogs or task managers and use cases like community hubs or education solutions. Additionally, the summary fails to address questions about XWiki's professional features and its use for project collaboration, indicating a lack of coverage of key aspects present in the original text.

Score_breakdown:

Alignment: 0.6

Coverage: 0.6

Input_language: en

Summary_language: en

Model: AI.Models.mistral-nemo_12b_Q4

File: summ_001_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary includes contradictions regarding the use of panels and extra information about customization not present in the original text. Additionally, the summary fails to answer a question about email notifications that the original text can address.

Score_breakdown:

Alignment: 0.6666666666666666

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 multiple contradictions regarding user rights and default settings compared to the original text. Additionally, the summary includes extra information about implied rights that are not mentioned in the original text. Furthermore, there are questions that the original text can answer but the summary cannot, indicating a lack of completeness and accuracy in the summary.

Score_breakdown:

Alignment: 0.7894736842105263

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.07142857142857142

Reason: The score is 0.07 because the summary includes a significant amount of extra information not present in the original text, which indicates a poor alignment with the original content. Additionally, the summary fails to answer questions that the original text can address, further reducing its effectiveness and relevance.

Score_breakdown:

Alignment: 0.07142857142857142

Coverage: 0.4

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 with the original text, such as misrepresenting the process of exporting and reimporting wiki pages and recommending the wrong downgrade method. Additionally, the summary includes extra information not present in the original text, like server restart instructions and Solr file editing, which can mislead the reader. Furthermore, the summary fails to address several questions that the original text can answer, indicating a lack of comprehensiveness and accuracy.

Score_breakdown:

Alignment: 0.25

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.7777777777777778

Reason: The score is 0.78 because the summary includes some contradictions, such as the ability to modify existing feeds through the Main.WebRss page, which is not mentioned in the original text. Additionally, the summary adds extra information by specifying a particular browser, Firefox, which is not mentioned in the original text. Despite these issues, the summary captures the main points reasonably well.

Score_breakdown:

Alignment: 0.7777777777777778

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_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 users learning to create, edit, and view page history in XWiki, and it lacks clarity on

whether users can learn about editing a page from the provided links. Additionally, it mentions setting permissions without explicitly stating it in the original text, and it implies developers can explore certain resources, which is not specified.

Score_breakdown:

Alignment: 0.5714285714285714

Coverage: 0.8

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 address a question that the original text can answer. 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: 0.5555555555555556

Reason: The score is 0.56 because the summary contains multiple contradictions regarding configurations and user rights, includes extra information not present in the original text, and fails to answer a question that the original text can address. These issues significantly affect the accuracy and completeness of the summary.

Score_breakdown:

Alignment: 0.5555555555555556

Coverage: 0.8

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 included.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_010_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. Excellent summarization quality.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0
Input_language: en
Summary_language: en

Model: AI.Models.mistral2_large

File: summ_001_result.json

Score: 0.75
Reason: The score is 0.75 because the summary contains contradictions regarding the display of panels in XWiki, includes extra information not present in the original text, and fails to address a question about email notifications that the original text can answer.
Score_breakdown:
Alignment: 0.75
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 multiple contradictions and extra information not present in the original text, leading to significant inaccuracies. Additionally, the summary fails to answer specific questions that the original text can address, further diminishing its quality.
Score_breakdown:
Alignment: 0.6153846153846154
Coverage: 0.4
Input_language: en
Summary_language: en

File: summ_003_result.json

Score: 0.5384615384615384
Reason: The score is 0.54 because the summary contains a contradiction regarding how to change access rights, and it includes several pieces of extra information not mentioned in the original text, such as creating user groups, specific documentation, and details about setting rights for pages and further reading.
Score_breakdown:
Alignment: 0.5384615384615384
Coverage: 1.0
Input_language: en
Summary_language: en

File: summ_004_result.json

Score: 0.4
Reason: The score is 0.40 because the summary contains several contradictions with the original text, such as incorrect methods for upgrading and incorrect descriptions of processes like the Distribution Wizard and Solr initialization. Additionally, the summary includes extra information not present in the original text and fails to answer several questions that the original text addresses. These issues significantly reduce the quality and accuracy of the summary.
Score_breakdown:

Alignment: 0.6
Coverage: 0.4
Input_language: en
Summary_language: en

File: summ_005_result.json

Score: 0.875
Reason: The score is 0.88 because the summary includes extra information about formats like Atom or RDF that were not mentioned in the original text. However, there are no contradictions, indicating a generally accurate summarization.
Score_breakdown:
Alignment: 0.875
Coverage: 1.0
Input_language: en
Summary_language: en

File: summ_006_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 details about users learning specific tasks on XWiki and a step in a guide. Additionally, there is a question the original text can answer that the summary cannot, indicating some missing information.
Score_breakdown:
Alignment: 0.75
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 includes extra information not present in the original text, such as a 'Version Control page on XWiki.org' and details about changing the logo and panels as the next step in the guide. However, there are no contradictions, which indicates a generally accurate representation of the original content.
Score_breakdown:
Alignment: 0.7777777777777778
Coverage: 1.0
Input_language: en
Summary_language: en

File: summ_008_result.json

Score: 0.25
Reason: The score is 0.25 because the summary includes a significant amount of extra information that is not mentioned in the original text, indicating a lack of precision and relevance in the summarization.
Score_breakdown:
Alignment: 0.25
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: 0.8

Reason: The score is 0.80 because the summary includes extra information about examples of applications like blogs or task managers that 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.mixtral-8x22b

File: summ_001_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. These issues indicate that the summary is not fully aligned with the original content, affecting its quality.

Score_breakdown:

Alignment: 0.75

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.9090909090909091

Reason: The score is 0.91 because the summary includes extra information not present in the original text, specifically a tabular view summarizing rights and their details. However, there are no contradictions, indicating a generally high-quality summary.

Score_breakdown:

Alignment: 0.9090909090909091

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.65

Reason: The score is 0.65 because the summary includes several pieces of extra information not found in the original text, such as details about XWiki being a web development platform, specific processes for editing access rights, and additional information about permissions and administration. This indicates that while the summary captures some key points, it introduces content not present in the original, affecting its accuracy and relevance.

Score_breakdown:

Alignment: 0.65

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.5555555555555556

Reason: The score is 0.56 because the summary contains a contradiction regarding the recommended upgrade method, includes extra details not present in the original text, and fails to address a question that the original text can answer. These issues indicate a moderate level of deviation from the original content.

Score_breakdown:

Alignment: 0.5555555555555556

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.8

Reason: The score is 0.80 because the summary includes a contradiction about user capabilities in creating RSS pages and adds extra information about XWiki being open-source, which is not in the original text. Despite these issues, the summary is mostly accurate and informative.

Score_breakdown:

Alignment: 0.8

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.8333333333333334

Reason: The score is 0.83 because the summary includes extra information not present in the original text, such as specific actions like creating a page, editing a page, and accessing page history, and a section titled 'Step 1: Creating a page.' However, there are no contradictions, indicating a generally accurate summary with some additional details.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_007_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 a 'Version Control page on XWiki.org' and learning to change the logo and panels. Additionally, the summary fails to answer a question that the original text can, specifically whether all content added to the wiki is saved over time.

Score_breakdown:

Alignment: 0.75

Coverage: 0.8

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, such as predefined configurations, screenshots, visiting XWiki.org, and creating a basic application. Additionally, the summary fails to answer questions that the original text can, such as whether XWiki allows admins to create new user accounts and if there is a configuration for admin-only content viewing and editing. Despite no contradictions, these issues impact the summary's completeness and accuracy.

Score_breakdown:

Alignment: 0.6363636363636364

Coverage: 0.6

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. Excellent summarization quality.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_010_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

Model: AI.Models.phi3_medium-128k_14b_Q4

File: summ_001_result.json

Score: 0

Reason: The score is 0.00 because the summary fails to address several key questions that the original text can answer, indicating a lack of coverage and completeness in the summarization.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: unknown

File: summ_002_result.json

Score: 0

Reason: The score is 0.00 because the summary fails to answer several questions that the original text can address, indicating a lack of coverage and completeness in the summary.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.14285714285714285

Reason: The score is 0.14 because the summary includes a significant amount of extra information not present in the original text, and it fails to answer several questions that the original text can address. This indicates a lack of alignment with the original content, leading to a low score.

Score_breakdown:

Alignment: 0.14285714285714285

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.0

Reason: The score is 0.00 because the summary includes extra information not present in the original text and fails to address several questions that the original text can answer, indicating a poor alignment with the original content.

Score_breakdown:

Alignment: 0.25

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0

Reason: The score is 0.00 because the summary fails to address key questions that the original text can answer, indicating a lack of coverage and completeness.

Score_breakdown:

Alignment: 0

Coverage: 0.0
Input_language: en
Summary_language: unknown

File: summ_006_result.json

Score: 0
Reason: The score is 0.00 because the summary fails to address key questions that the original text can answer, indicating a lack of coverage and completeness in the summary.
Score_breakdown:
Alignment: 0
Coverage: 0.0
Input_language: en
Summary_language: unknown

File: summ_007_result.json

Score: 0
Reason: The score is 0.00 because the summary fails to address key questions that the original text can answer, indicating a lack of comprehensiveness and coverage.
Score_breakdown:
Alignment: 0
Coverage: 0.0
Input_language: en
Summary_language: unknown

File: summ_008_result.json

Score: 0.0
Reason: The score is 0.00 because the summary contains multiple contradictions and introduces extra information not present in the original text, leading to a significant deviation from the original content. Additionally, the summary fails to answer questions that the original text can address.
Score_breakdown:
Alignment: 0.0
Coverage: 0.8
Input_language: en
Summary_language: en

File: summ_009_result.json

Score: 0
Reason: The score is 0.00 because the summary fails to answer multiple questions that the original text can address, indicating it lacks essential information and completeness.
Score_breakdown:
Alignment: 0
Coverage: 0.0
Input_language: en
Summary_language: en

File: summ_010_result.json

Score: 0

Reason: The score is 0.00 because the summary fails to answer key questions that the original text addresses, indicating a lack of coverage and completeness.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: unknown

Model: AI.Models.phi3_mini-128k_4b_Q4

File: summ_001_result.json

Score: 0

Reason: The score is 0.00 because the summary fails to answer several questions that the original text can answer, indicating it lacks essential information and context.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0

Reason: The score is 0.00 because the summary fails to address several key questions that the original text can answer, indicating a lack of coverage and completeness in the summarization.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.0

Reason: The score is 0.00 because the summary includes a significant amount of extra information not present in the original text, indicating a poor alignment between the summary and the original content.

Score_breakdown:

Alignment: 0.0

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.0

Reason: The score is 0.00 because the summary includes extra information not present in the original text and fails to answer several questions that the original text can address. This indicates a significant deviation from the original content, leading to a poor summarization quality.

Score_breakdown:

Alignment: 0.0

Coverage: 0.0

Input_language: en
Summary_language: en

File: summ_005_result.json

Score: 0

Reason: The score is 0.00 because the summary fails to answer several key questions that the original text addresses, indicating a lack of coverage and completeness in the summary.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.0625

Reason: The score is 0.06 because the summary contains significant contradictions and introduces a large amount of extra information not present in the original text, indicating a poor alignment with the original content.

Score_breakdown:

Alignment: 0.0625

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0

Reason: The score is 0.00 because the summary fails to address several questions that the original text can answer, indicating a lack of coverage and completeness.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.45

Reason: The score is 0.45 because the summary contains multiple contradictions regarding user permissions and settings, such as incorrect claims about who can view, comment, and edit in various modes like 'Editable', 'Closed', and 'Hidden'. Additionally, the summary includes extra information about settings not mentioned in the original text, such as 'Viewable Closed (Light)'. Furthermore, the summary fails to address a specific question about overriding global rights at the page level, which the original text can answer.

Score_breakdown:

Alignment: 0.45

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0

Reason: The score is 0.00 because the summary fails to address several key questions that the original text can answer, indicating a lack of coverage and comprehensiveness.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0

Reason: The score is 0.00 because the summary fails to answer key questions that the original text addresses, indicating a lack of coverage of important information.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: en

Model: AI.Models.qwen2_7b_Q4

File: summ_001_result.json

Score: 0.3

Reason: The score is 0.30 because the summary contains multiple contradictions and extra information not present in the original text, and it fails to answer a question that the original text can address.

Score_breakdown:

Alignment: 0.3

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 several contradictions with the original text, such as incorrect claims about editing rights, programming rights, and the hierarchy of rights. Additionally, it includes extra information not mentioned in the original text and fails to answer specific questions that the original text can address.

Score_breakdown:

Alignment: 0.75

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.09090909090909091

Reason: The score is 0.09 because the summary contains significant contradictions with the original text, introduces a large amount of extra information not present in the original text, and fails to answer several questions that the original text can address. These issues indicate a low-quality summarization.

Score_breakdown:

Alignment: 0.09090909090909091

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.0

Reason: The score is 0.00 because the summary includes a significant amount of extra information not present in the original text and fails to address questions that the original text can answer.

Score_breakdown:

Alignment: 0.09090909090909091

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains contradictions regarding the source of the feed for modified pages and includes extra information about Atom or RDF that is not in the original text. Additionally, the summary fails to address questions about creating new RSS feeds and receiving information via a browser's RSS notification button, which the original text can answer.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.375

Reason: The score is 0.38 because the summary contains multiple inaccuracies and adds information not present in the original text. It incorrectly describes the content of the Getting Started guide and the availability of applications for developers. Additionally, it introduces several details that are not mentioned in the original text, such as the need for contacting an administrator for login credentials and steps for creating pages. Furthermore, the summary fails to answer a question that the original text can address, indicating a lack of completeness and accuracy.

Score_breakdown:

Alignment: 0.375

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.8

Reason: The score is 0.80 because the summary contains a minor contradiction regarding the location and function of the 'History' tab and includes extra information about a 'Version Control page' not present in the original text. Despite these issues, the summary largely aligns with the original text, indicating a generally high-quality summarization.

Score_breakdown:

Alignment: 0.8

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 contains a contradiction regarding user permissions and includes extra details not present in the original text, such as specific override options, visual examples, and next steps in application creation.

Score_breakdown:

Alignment: 0.75

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 includes extra information not present in the original text, such as mentioning XWiki or specific wiki software, and it fails to answer a question about modifying a page by clicking an 'edit' button that the original text can answer. However, there are no contradictions, indicating a generally accurate representation of the original text.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_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 portals integrating external data sources, competitive intelligence platforms, public debates, or collaborative course creation in education. Additionally, the summary fails to answer whether XWiki is a second-generation wiki, which the original text can answer.

Score_breakdown:

Alignment: 0.75

Coverage: 0.8

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 accurately extracts the key information from the Input, maintaining relevance and coherence. However, it does not match the Expected Output format, which affects the alignment with the intended message.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output effectively addresses the Input prompt by incorporating all keywords into a coherent sentence. It maintains logical flow and consistency with the Expected Output, differing only in the choice of 'frolicked' instead of 'enjoyed exploring', which does not significantly impact fluency or coherence.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output addresses the Input prompt by explaining photosynthesis in a simple way suitable for ages 8-10, similar to the Expected Output. It maintains coherence and fluency with a logical flow and clear language. The quality is slightly more playful than the Expected Output but still meets the prompt's standards.

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 points in the Input, such as noise-cancelling, 20-hour battery life, and comfortable fit, targeting tech-savvy professionals. It aligns well with the Expected Output, maintaining key ideas and themes. The coherence and fluency are strong, with a logical flow and smooth language, though slightly more descriptive than the Expected Output.

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 creative story about a space adventure, aligning well with the Input prompt. It is coherent, with a logical flow and consistent narrative about Captain Vega and her crew exploring a new planet, Lumina. The story is fluent, with correct grammar and smooth readability. It matches the Expected Output by featuring astronauts exploring a new planet, fulfilling all key elements and quality criteria.

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 captures the beauty of nature, aligning perfectly with the Input prompt and the Expected Output's intent 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 a well-structured news article that directly addresses the Input prompt by detailing a recent scientific discovery in quantum computing. It aligns with the Expected Output by being informative and summarizing the breakthrough effectively. The article is coherent, logically flowing from the introduction of the discovery to its implications, and is fluent with clear grammatical structure and readability.

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, directly addressing the Input prompt by providing detailed information on attractions, accommodations, and activities. It is coherent, with a logical flow from introduction to conclusion, and maintains fluency with correct grammar and natural language. The content aligns well with the Expected Output, meeting the intended goals of highlighting key aspects of a popular tourist destination.

Prompt_language: fr
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, dates, location, and features, aligning with the Input. However, it combines the venue and address and does not explicitly separate the start and end dates or mention 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: 0.8

Reason: The Actual Output is relevant to the Input, addressing the complexity of the brain. It is coherent and aligns well with the Expected Output, though it is slightly less detailed, missing 'coordinating' and 'subjective experiences.' The fluency is consistent with the Expected Output.

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, maintaining relevance and coherence. However, it does not match the JSON format of the Expected Output, which affects alignment with the intended message.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output addresses the Input by forming a coherent sentence with the given keywords. It maintains a logical flow and consistency with the Expected Output, though it uses 'frolicked joyfully' instead of 'enjoyed exploring'. The language is fluent and relevant to the core subject.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input and Expected Output, explaining photosynthesis in a way that is engaging for a younger audience. It is coherent and maintains consistency with the main topic. The language is fluent, grammatically correct, and natural, though it introduces the concept as a 'magic trick,' which is slightly different from the Expected Output's straightforward explanation.

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 prompt, addressing all key features and the target audience. It is coherent and fluent, with logical flow and correct language use. While it closely aligns with the Expected Output, it uses slightly different phrasing but maintains the same intent and key elements.

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 features astronauts exploring a new planet, consistent with the Expected Output. The story is coherent, fluent, and maintains a logical flow, with only minor deviations in the specific details of exploration.

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 a poem about 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 a well-structured news article that clearly addresses the Input prompt by detailing a recent scientific discovery about brain plasticity. It aligns with the Expected Output by providing an informative summary of the finding. The article is coherent, with a logical flow of information, and is fluent, with accurate grammar and natural language use.

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, aligning perfectly with the Input prompt. It covers top attractions, accommodations, and activities, matching the Expected Output's criteria for an informative guide. The content is coherent, logically structured, and grammatically fluent.

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 and presents the event details, addressing the Input correctly. It is coherent and fluent, with no grammatical errors. However, it does not separate the start and end dates or explicitly mention 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 in a technical manner. It is coherent and maintains a logical flow, expanding on neural architecture and integration. The fluency is strong, with correct grammar and natural language. It closely matches the Expected Output, with minor differences in phrasing but similar technical depth.

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, affecting 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, maintains coherence with the Expected Output, and is fluent with proper grammar and syntax. The only difference is a slight variation in word order, which does not affect the overall meaning.

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 'super power' and comparing it to how kids need food for energy. It maintains coherence with the Expected Output by explaining the process of plants making their own food using sunlight. The fluency is high with clear grammar and syntax, though it adds a playful tone 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 maintains coherence with a logical flow and aligns closely with the Expected Output structure. The language is fluent and grammatically correct, matching the quality of the Expected Output.

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 well with the Input prompt. It is coherent, fluent, and engaging. However, it focuses on Europa rather than a new planet as mentioned in the Expected Output, which slightly misaligns with the expected key points.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input by providing a poem about nature, aligning perfectly with the Expected Output's theme of describing the wonders of the natural world. The poem is coherent, with a logical flow and consistent imagery throughout, matching the Expected Output's intent. It is also fluent, with correct grammar and natural language use, comparable to the Expected Output's fluency level.

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 generating a news article about a recent scientific discovery, aligning well with the Expected Output's requirement for an informative summary. It maintains coherence and fluency throughout, with logical flow and readability. The only deviation is the fictional nature of the discovery, which slightly affects relevance.

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 covers key attractions, accommodations, and activities, aligning with the Expected Output. The information is coherent, fluent, and logically structured, meeting all evaluation criteria.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_009_result.json

Score: 0.9

Reason: The Actual Output accurately extracts and presents the event details from the Input, aligning closely with the Expected Output. It includes all relevant information such as event name, dates,

location, and target audience. The format differs slightly, but the content is coherent and fluent.

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, using technical language suitable for the audience. It maintains coherence and fluency with logical connections and natural language. It aligns well with the Expected Output, covering key points like physiological processes and cognitive functions, though it slightly differs in phrasing.

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 correctly extracts and presents the key information from the Input, maintaining relevance and coherence. However, it slightly differs in format from the Expected Output by not using JSON structure.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.8

Reason: The Actual Output accurately reflects the Input keywords and maintains coherence and fluency. However, it slightly differs in structure from the Expected Output, focusing on the cat's feelings rather than its actions in the garden.

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 engaging language and relatable metaphors, directly addressing the Input prompt. It maintains coherence and fluency with a logical flow and correct grammar. While it adds creative elements not present in the Expected Output, it still meets the criteria of explaining photosynthesis in a way that is relevant and understandable for children.

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 is coherent and fluent, with logical flow and correct grammar. It closely matches the Expected Output in terms of quality and content, though it lacks the explicit mention of 'cutting-edge' and 'premium' descriptors.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.8

Reason: The Actual Output addresses the key elements of the Input prompt by presenting a space adventure with exploration and discovery, aligning with the Expected Output's theme of astronauts exploring. The story is coherent and fluent, with a logical flow and consistent narrative. However, the focus on a duo rather than a group of astronauts slightly deviates 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 aligns well with the Input prompt by vividly describing the beauty of nature, matching 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: 0.9

Reason: The Actual Output is highly relevant to the Input, as it discusses a recent scientific discovery about a new material. It is coherent and logically structured, detailing the discovery, its implications, and potential applications, aligning well with the Expected Output of an informative article. The fluency is smooth and readable, closely matching the quality expected.

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 Rome, which aligns well with the Input request for a travel guide to a popular tourist destination. It covers top attractions and activities, matching the Expected Output's focus on attractions and activities. The guide is coherent and fluent, with a logical day-by-day itinerary and smooth sentence structure. However, it lacks specific mention of accommodations, which slightly deviates from the Expected Output.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output captures the main event details such as event name, date, and location, aligning with the Input. However, it lacks the separation of start and end dates and does not explicitly mention the target audience as 'professionals in the tech industry,' affecting coherence with 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 addresses the main topic by describing the brain's complexity and functions. It covers similar key points to the Expected Output, such as physiological functions and cognitive abilities, though it lacks mention of consciousness and subjective experiences. The ideas flow logically and the language is grammatically correct and natural.

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 correctly extracts the name, date of birth, and address from the Input, aligning with the Expected Output in terms of content. However, it uses a different format with bullet points and bold text instead of 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 addresses the main topics and themes from the Input, maintaining relevance. It is coherent with a logical flow similar to the Expected Output, although the verb 'frolicked' differs slightly from 'enjoyed exploring'. The sentence structure, grammar, and readability are fluent and consistent with 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, using simple language and relatable metaphors like 'little food factories.' It maintains coherence and

fluency throughout. While it introduces additional elements like 'air and water,' which aren't in the Expected Output, it still aligns well with the Input prompt and 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 topics in the Input, focusing on noise-cancelling, battery life, and comfort, relevant to tech-savvy professionals. It aligns well with the Expected Output, covering key features and benefits. The flow is logical and coherent, with clear and natural language. Minor differences in phrasing prevent a perfect score.

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 well with the Input prompt. It features a space crew on a mission, addressing the topic effectively. However, it diverges from the Expected Output, which specifies a story about astronauts exploring a new planet, whereas the Actual Output focuses on a nebula and a mythical plant. Despite this, the story is coherent, fluent, and engaging.

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 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: 0.9

Reason: The Actual Output effectively addresses the Input by detailing a recent scientific discovery about ancient sea creatures and their potential impact on autoimmune disease treatments, aligning well with the Expected Output of summarizing a groundbreaking finding. The language is fluent and the article maintains logical consistency. However, it could slightly improve by explicitly stating the discovery's recentness.

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 Amsterdam, directly addressing the Input prompt. It includes detailed sections on attractions, accommodations, and activities, aligning perfectly with the Expected Output. The language is fluent and grammatically correct, maintaining a logical flow throughout.

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 and presents all relevant event details from the Input, aligning well with the Expected Output. It includes the event name, dates, location, activities, and target audience. The only minor deviation is the format of dates and the additional 'Activities' section, which does not affect coherence or 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 effectively adapts the Input for a technical audience by using precise language like 'biological network' and 'expression of cognition and consciousness.' It maintains coherence and fluency similar to the Expected Output, though it slightly diverges by not explicitly mentioning 'subjective experiences.'

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 contains all key information correctly extracted from the Input, maintaining relevance and coherence. However, the format differs from the Expected Output, which affects the fluency assessment.

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, addressing all keywords. It is coherent with a logical flow, though slightly more detailed than the Expected Output. The fluency is high, matching the linguistic 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 by using relatable language like 'superheroes' and 'magic process.' It maintains coherence and fluency, with a logical flow and correct grammar. While it adds a creative touch not present in the Expected Output, it still conveys the key idea of plants using sunlight to make energy.

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 by describing the product's features and targeting tech-savvy professionals. It is coherent, fluent, and grammatically correct. While it slightly differs in wording from the Expected Output, it maintains the same quality and relevance, effectively highlighting noise-cancelling, battery life, and comfort.

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 space adventure, directly addressing the Input prompt. It aligns with the Expected Output by featuring astronauts exploring a new planet, maintaining consistency and grammatical correctness throughout.

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 effectively captures the beauty of nature, aligning perfectly with the Input prompt and the Expected Output's objective 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 a well-written news article that directly addresses the Input prompt by detailing a recent scientific discovery. It is coherent and logically structured, providing specific details about the discovery of *Homo luzonensis*, which aligns with the Expected Output of summarizing a groundbreaking scientific finding. The language is fluent and grammatically correct, matching 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 a comprehensive travel guide for Santorini, directly addressing the Input prompt by providing detailed information on attractions, accommodations, and activities, aligning perfectly with the Expected Output. The content is coherent, fluent, and logically organized, covering all necessary aspects of a travel guide.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output correctly identifies the event name, dates, and location, maintaining logical flow and fluency. However, it combines start and end dates and misses the target audience detail 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 uses technical language and addresses the key elements of the Input prompt, such as brain functions and sensory processing. It maintains coherence and fluency, with logical connections between ideas. While it closely aligns with the Expected Output, it slightly differs in phrasing but still meets the anticipated quality for a technical audience.

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 accurately extracts the key information from the Input, maintaining relevance and coherence. However, it does not match the Expected Output format, which affects alignment with the intended message.

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, coherent, and fluent. It includes all keywords and maintains logical flow. It aligns with the Expected Output, though it adds extra details about the cat lounging and soaking up the sunshine.

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 with the Input prompt. It maintains coherence and logical flow, similar to the Expected Output, while being more engaging. The fluency is high, with clear grammar and readability, though it adds extra details 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 aligns well with the Expected Output, addressing the product's features and target audience. It maintains logical flow and consistency, with minor differences in wording. The fluency is excellent, matching the grammatical correctness and smoothness of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.7

Reason: The story is relevant to the input prompt as it describes a space adventure, but it diverges from the expected output by focusing on a nebula rather than a new planet. The narrative is coherent and fluent, with logical progression and high language quality, but it lacks alignment with the specific focus on exploring a new planet as expected.

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 nature, aligning perfectly with the Input prompt and 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 directly addresses the Input by presenting a news article on a recent scientific discovery. It maintains logical flow and consistency, detailing the discovery of an ancient tree species, which aligns with the Expected Output's requirement for a summary of a groundbreaking finding. The article is fluent, grammatically correct, and smoothly written, matching the Expected Output's style.

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 Tokyo, Japan, aligning perfectly with the Input prompt. It covers attractions, accommodations, and activities, matching the Expected Output's criteria. The content is relevant, coherent, and fluent, providing detailed and logically structured information about Tokyo as a tourist destination.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.7

Reason: The Actual Output accurately extracts the event name, dates, and location, aligning well with the Input. However, it lacks details on the target audience and separates the venue and address differently than the Expected Output, 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 aligns well with the Input prompt and Expected Output by using technical language and addressing the same topics. It maintains coherence and fluency, though it focuses more on the cerebral cortex rather than the entire brain, which is a slight deviation from the Expected Output's broader scope.

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 correctly extracts the key information from the Input and presents it clearly. 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: 0.8

Reason: The Actual Output is relevant to the Input prompt, maintaining coherence and fluency. However, it slightly differs in style and structure from the Expected Output, which emphasizes the cat's enjoyment and exploration.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input, explaining photosynthesis in a child-friendly manner. It maintains coherence with a logical flow, expanding on the Expected Output by adding engaging details about the process and its benefits. The language is fluent and suitable for the target age group, though slightly more detailed than 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 highlighting the product's features and target audience. It maintains coherence and fluency with a logical flow and smooth language. While it closely matches the Expected Output, it slightly lacks the introductory tone 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. However, it deviates from the Expected Output, which specifies exploration of a new planet, not a cosmic structure. The Actual Output is coherent and fluent but lacks direct alignment with the Expected Output's focus on planetary exploration.

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 prompt and Expected Output, as it beautifully describes the wonders of nature. The poem is coherent, with well-structured stanzas that logically flow from one idea to the next, capturing various elements of the natural world. The language is fluent, with excellent grammar and readability, matching the Expected Output's requirement for a descriptive poem about nature.

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 by providing a detailed news article about a recent scientific discovery, aligning with the Expected Output's requirement for an informative summary. It maintains coherence by logically presenting the discovery's context, significance, and implications. The article is fluent, with correct grammar and readability, 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 is a detailed travel guide for Paris, directly addressing the Input prompt. It is coherent, with a logical flow covering attractions, transportation, and tips, matching the Expected Output's informative nature. The fluency is high with clear and structured sentences.

Prompt_language: fr

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, venue, and address, aligning well with the Expected Output. However, it lacks the 'target_audience' key and separates dates into a single entry instead of 'start_date' and 'end_date'. The coherence and fluency are strong, with logical structure and correct grammar.

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 incorporating specific terminology like 'neurons,' 'glia,' and 'neural circuits,' which aligns well with the Expected Output. It maintains coherence and fluency, though it includes slightly more detail than necessary, such as ongoing research, which was not 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: 0.8

Reason: The Actual Output accurately extracts the key information from the Input, maintaining relevance. However, it does not match the JSON format of the Expected Output, affecting coherence

and fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.8

Reason: The Actual Output addresses the Input prompt and is relevant to the Expected Output. It is coherent and maintains logical flow, but lacks the word 'exploring' which is present in the Expected Output. The sentence is fluent and grammatically correct.

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, maintaining relevance by addressing the main topic of energy conversion from sunlight. It is coherent and logically flows, similar to the Expected Output, though it slightly lacks the explicit mention of 'making food'. The fluency is appropriate for the age group, with clear and engaging language.

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 target audience. It maintains coherence and logical flow similar to the Expected Output, with minor differences in wording. Sentence structure and grammar are fluent and comparable to the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.5

Reason: The Actual Output is a creative space adventure story, aligning with the Input prompt. However, it diverges from the Expected Output by focusing on fictional characters and elements rather than astronauts exploring a new planet. The coherence, fluency, and context are well-maintained, but the thematic focus differs 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 poem beautifully captures the beauty of nature, aligning perfectly with the prompt. It is coherent with vivid imagery and smooth transitions between stanzas. The language is fluent and

error-free, and it matches the expected output by 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 directly addresses the Input by presenting a news article about a recent scientific discovery. It is coherent, with a logical flow and consistent ideas, matching the Expected Output of summarizing a groundbreaking scientific finding. The fluency is excellent, with clear sentence structure and grammar.

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 by detailing top attractions, accommodations, and activities. It is coherent, logically structured, and maintains a consistent flow throughout. The language is fluent, grammatically correct, and matches the natural language use expected.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output captures the main event details like name, date, location, and activities, aligning well with the Input. However, it differs in format and lacks explicit separation of start and end dates, and the venue and address are combined. The target audience is correctly identified.

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 and Expected Output, addressing the complexity and functions of the brain. It is coherent, with a logical flow and detailed explanation of brain functions, similar to the Expected Output. The language is fluent and technical, suitable for the intended audience, though slightly more detailed than necessary.

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 is relevant and coherent with the Input and Expected Output, extracting the correct information. However, it lacks JSON formatting, affecting 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 and coherent, directly addressing the Input prompt with a logical flow. It is fluent with correct grammar and syntax. The main difference from the Expected Output is the use of 'enjoys frolicking' instead of 'enjoyed exploring,' which slightly affects alignment but maintains 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 by using relatable language and metaphors like 'magic trick' and 'special ingredient,' which aligns well with the Expected Output. It maintains relevance and coherence, though it adds a bit more playful language than 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 themes of the Input, including noise-cancelling, battery life, and comfort, targeting tech-savvy professionals. It is coherent and fluent, with logical flow and correct language. It closely aligns with the Expected Output, covering key points like noise-cancelling, 20-hour battery life, and ergonomic fit, though it uses slightly different phrasing.

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 a creative space adventure story that aligns with the Input prompt. It is coherent and fluent, with a logical flow and correct grammar. However, it diverges from the Expected Output by focusing on a nebula and alien civilization rather than a new planet exploration.

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 of a poem about 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, focusing on a recent scientific discovery about converting CO2 into fuel. It is coherent, with a logical flow and consistent ideas, aligning well with the Expected Output of summarizing a groundbreaking finding. The language is fluent and readable, meeting the Expected Output's standard.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output thoroughly addresses the main topic of creating a travel guide for Paris, a popular tourist destination. It is coherent, logically structured, and maintains fluency with correct grammar and readability, aligning well with the Expected Output. The guide provides detailed information on attractions, accommodations, and activities, matching the specificity required by the Input.

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, dates, location, and details, aligning well with the Input. However, it lacks the separation of start and end dates and does not specify the target audience as distinctly as 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, using technical terminology such as 'cerebrum' and 'neurobiological structure.' It is coherent and fluent with logical flow and advanced vocabulary. It aligns closely with the Expected Output, addressing key points like regulation, sensory processing, and cognitive functions, though it slightly differs in phrasing.

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: 0.9

Reason: The Actual Output correctly extracts the key information from the Input and presents it clearly. 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 by incorporating all keywords into a coherent sentence. It maintains logical flow and consistency with the Expected Output, differing only by using 'frolicking in the vibrant garden' instead of 'exploring the garden,' which slightly enhances descriptiveness but remains aligned with the intended meaning. The sentence is fluent, grammatically correct, and readable.

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 addresses the Input by explaining photosynthesis in a fun and engaging way for children, aligning with the Expected Output's topic. It maintains coherence by logically explaining the process and uses child-friendly language, enhancing fluency. It slightly exceeds the Expected Output by adding relatable analogies, making it more appealing to 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 all features and the target audience. It maintains coherence with a logical flow and aligns closely with the Expected Output's intent. The language is fluent with minor typographical errors, such as 'cutting-e edge' instead of 'cutting-edge'.

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 story about a space adventure. The story is coherent and fluent, with logical progression and smooth language. However, it diverges from the Expected Output by focusing on a single astronaut rather than a group 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 highly relevant to the Input prompt, effectively capturing the beauty of nature. It maintains coherence with a logical flow and consistent ideas, aligning well with the Expected Output. The poem is fluent, with varied sentence structures and grammatical accuracy, matching the quality expected.

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, discussing a recent scientific discovery in quantum computing. It is coherent and logically structured, covering the discovery, implications, and technical details, aligning well with the Expected Output of summarizing a groundbreaking finding. Minor grammatical errors like 'advancebrought' slightly affect fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 0.7

Reason: The Actual Output addresses the Input by creating a travel guide for Paris, aligning with the Expected Output's requirement for an informative guide. It is coherent and logically flows from the Input, providing relevant information about transportation. However, it lacks details on top attractions, accommodations, and activities, which are specifically mentioned in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.6

Reason: The Actual Output correctly identifies the event name, location, and address but lacks coherence in date formatting and omits the target audience. The Expected Output provides a more structured and complete representation.

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 and uses technical language appropriate for the audience, similar to the Expected Output. It maintains coherence and fluency, but slightly deviates by focusing on the cerebral cortex rather than the entire brain as in 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: 0.0
Reason: The Actual Output is incoherent, does not address the Input prompt, and is irrelevant to the Expected Output.
Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_002_result.json

Score: 0.0
Reason: The Actual Output is completely irrelevant to the Input prompt, which involves transforming specific keywords into a coherent sentence. The Actual Output is incoherent and unrelated to the Expected Output, which is a simple sentence about a cat in a 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 addresses the Input by explaining photosynthesis in a child-friendly manner. It maintains coherence by describing the process with relatable imagery and includes additional details like oxygen production, which is consistent with the Expected Output's focus on energy conversion. However, it is slightly more complex than necessary for the target age group, affecting fluency.
Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_004_result.json

Score: 0.0
Reason: The Actual Output is irrelevant to the Input prompt, lacking any mention of wireless Bluetooth headphones or their features. It is incoherent and unrelated to the Expected Output, which clearly describes the product and its attributes.
Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_005_result.json

Score: 0.0
Reason: The Actual Output does not address the Input prompt about a space adventure story and is incoherent with numerous unrelated topics. It does not align with the Expected Output of a story about

astronauts exploring a new planet.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 0.0

Reason: The Actual Output does not address the theme of nature's beauty from the Input, nor does it align with the Expected Output of a nature poem. It lacks coherence, fluency, and relevance to the prompt.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 0.0

Reason: The Actual Output is incoherent, irrelevant to the Input prompt, and lacks any logical flow or readability compared to the Expected Output of an informative article on a scientific discovery.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 0.8

Reason: The Actual Output provides a comprehensive travel guide for Paris, aligning well with the Input prompt. It includes detailed descriptions of attractions and activities, which matches the Expected Output. However, the inclusion of a hotel experience and an itinerary for an elderly couple, which were not part of the original prompt, slightly detracts from coherence and relevance.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.0

Reason: The Actual Output is incoherent and irrelevant to the Input prompt, failing to extract any event details. It does not align with the Expected Output or maintain fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.0

Reason: The Actual Output is irrelevant to the Input prompt and lacks coherence, as it does not address the topic of the human brain. It also fails in fluency, with disjointed ideas and poor grammatical structure, unlike the Expected Output which is clear and technical.

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 name and date of birth, and accurately breaks down the address into components, which aligns with the Expected Output. However, the address format differs slightly from the Expected Output, affecting the overall alignment.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output addresses all key points from the Input, maintaining relevance. It presents the main ideas consistently with the Expected Output, ensuring coherence. The grammatical structure and word choice are fluent and consistent with the Expected Output, with only slight variations in phrasing.

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 addresses the Input prompt by explaining photosynthesis in a way that is engaging for a younger audience. It aligns with the Expected Output by describing how plants use sunlight to create energy. The output is coherent and fluent, with added elements like 'magic' and 'pretty flowers' to enhance appeal to children, though slightly more detailed than 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, directly addressing the features and target audience. It is coherent and fluent, with ideas logically connected and easy to understand. Compared to the Expected Output, it includes all key elements such as noise-cancelling, battery life, and comfort, while adding more detail about the design and use cases, which aligns with the intent.

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 coherent and fluent story about astronauts exploring Io, one of Jupiter's moons, which aligns with the Input prompt for a space adventure. It maintains a logical flow and is free of grammatical errors. While it focuses on a moon rather than a new planet, it meets the key points and intentions of the Expected Output by detailing an exploration mission with astronauts.

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 effectively addresses the Input by 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 is highly relevant to the Input, presenting a recent scientific discovery in neuroscience. It is coherent, with a logical flow and consistent ideas, matching the Expected Output's requirement for an informative article on a groundbreaking finding. The language is fluent, grammatically correct, and natural, aligning well with the Expected Output.

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 Paris, aligning well with the Input prompt for a popular tourist destination. It includes sections on attractions, transportation, and local cuisine, which are relevant and coherent with the Expected Output. However, it lacks specific details on accommodations, which slightly affects completeness.

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 event details such as name, date, location, and activities, aligning well with the Input and Expected Output. However, it lacks specific JSON formatting and does not explicitly mention '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 prompt by adapting the text for a technical audience, aligning well with the Expected Output in terms of relevance and coherence. It maintains fluency and introduces technical terms like 'central processing unit' and 'neural networks.' However, it slightly diverges by adding more detail than the Expected Output, such as mentioning 'creativity, emotion, and motivation.'

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en