

LLM Evaluation Report

Executive Summary

Methodology

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

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

Find the full methodology document at:

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

Metrics

Answer Relevancy

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

Faithfulness

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

Contextual Precision

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

Contextual Recall

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

Custom Context Relevancy

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

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

Evaluation Criteria:

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

Model Information

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

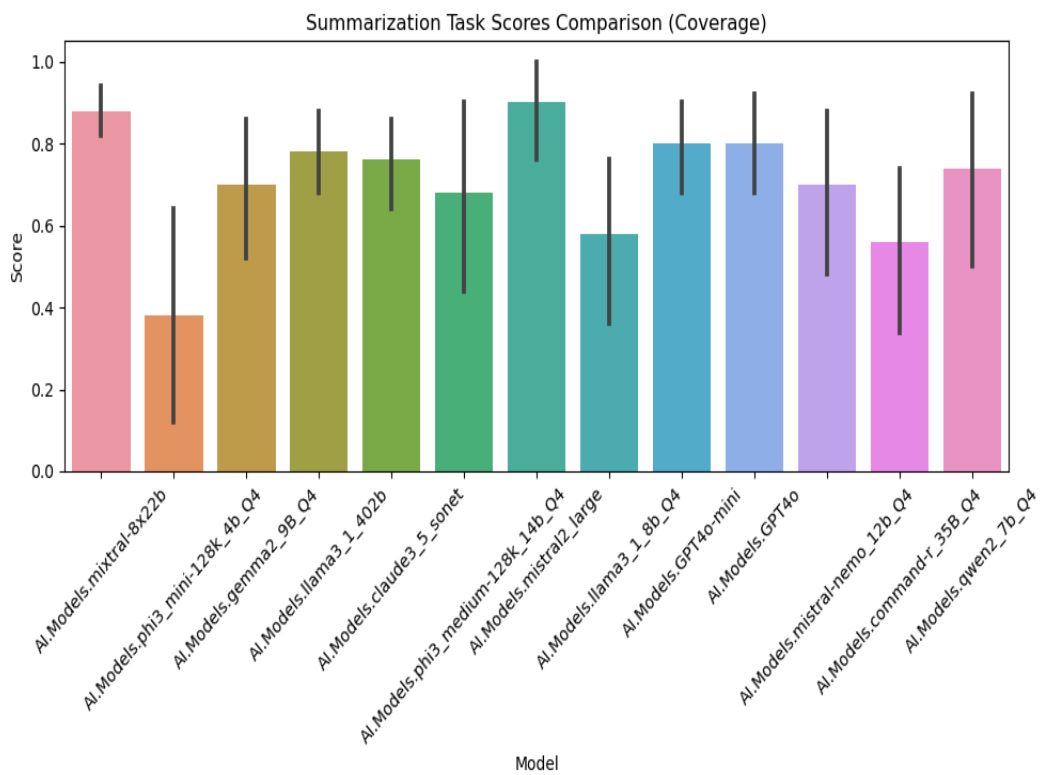
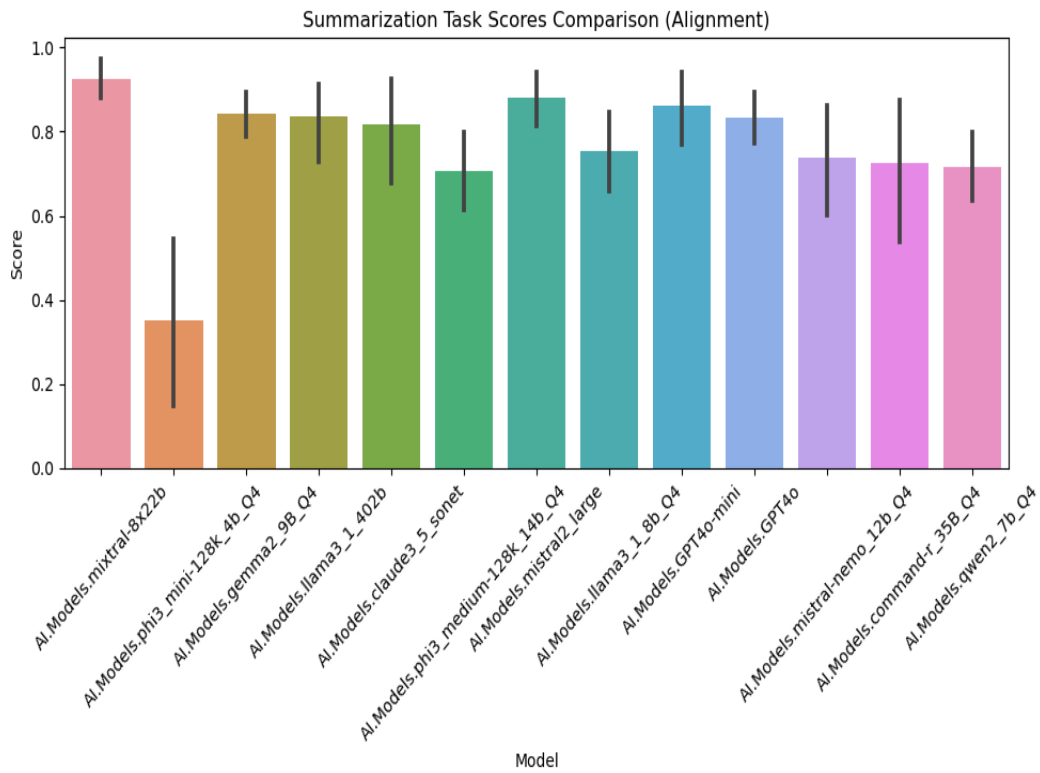
Setup Information

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

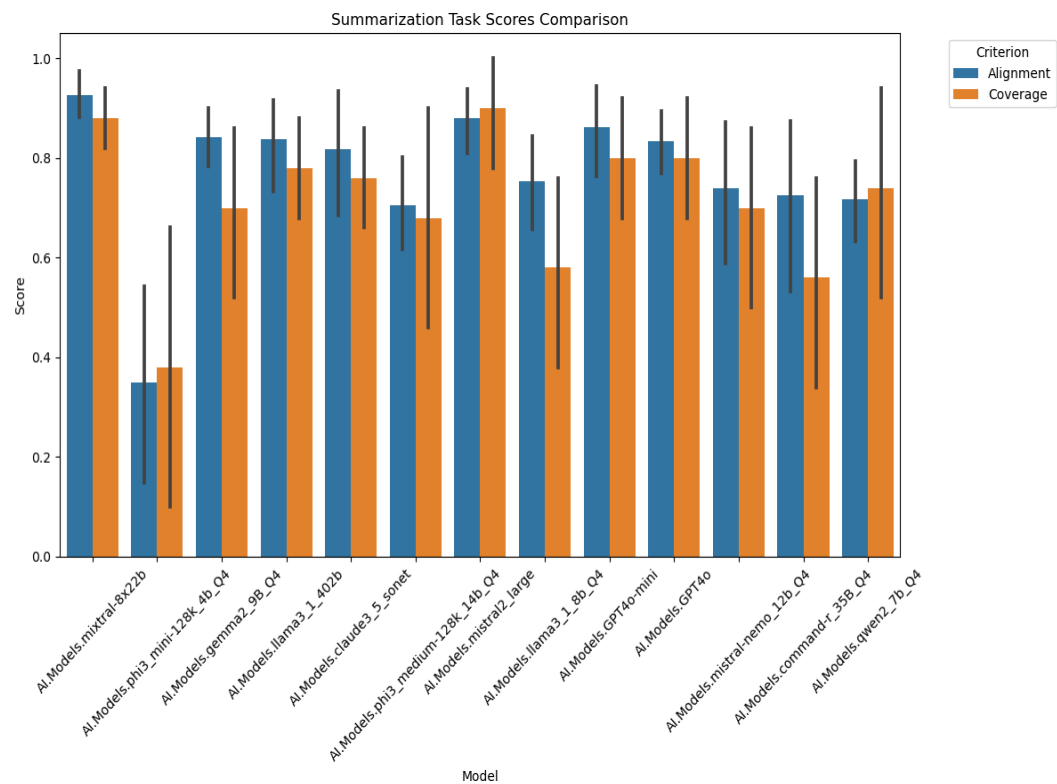
RAG-qa	AI.Models.qa_GPT4o	0.8	No
RAG-qa	AI.Models.qa_GPT4o-mini	0.8	No
RAG-qa	AI.Models.qa_claude3_5_sonnet	0.8	No
RAG-qa	AI.Models.qa_mistral2_large	0.8	No
RAG-qa	AI.Models.qa_llama3_1_402b	0.45	No
RAG-qa	AI.Models.qa_llama3_1_8b_Q4	0.45	Yes
RAG-qa	AI.Models.qa_mixtral-8x22b	0.8	No
RAG-qa	AI.Models.qa_mistral-nemo_12b_Q4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

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

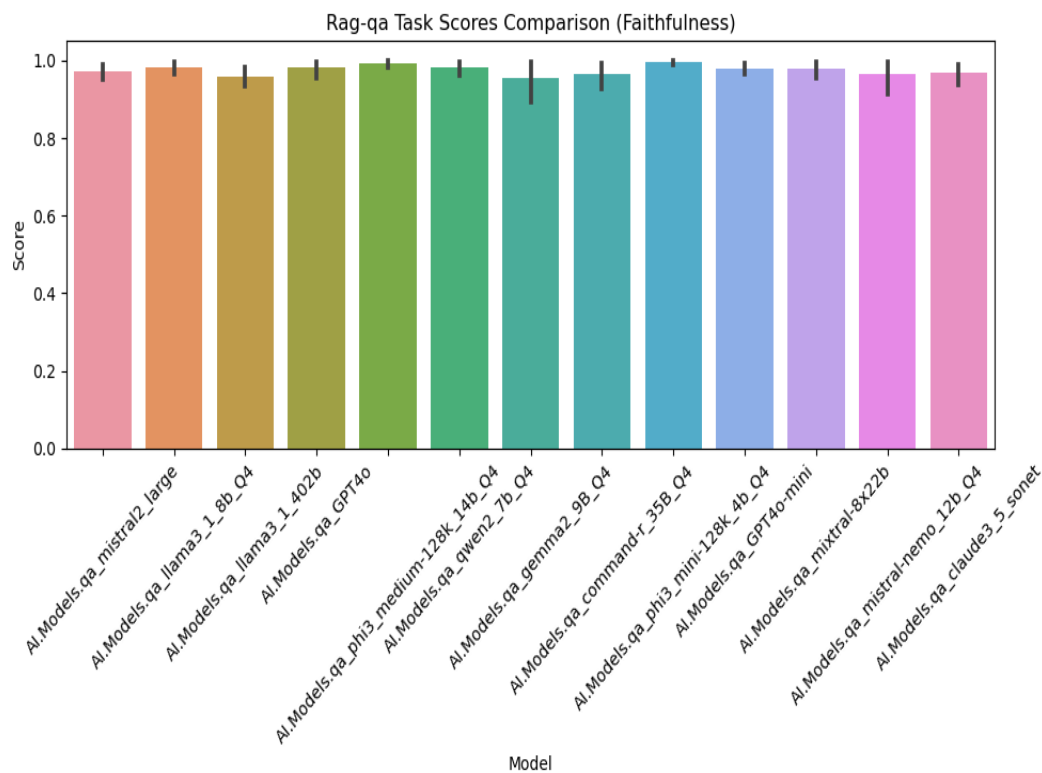
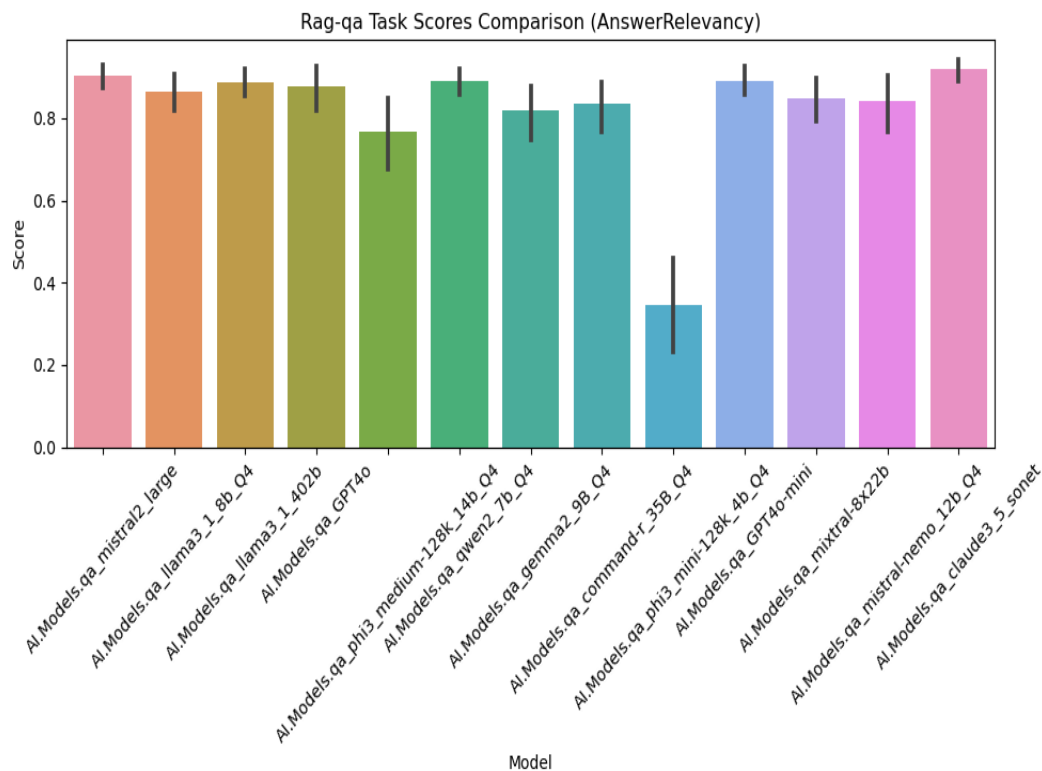
Summarization Task Results

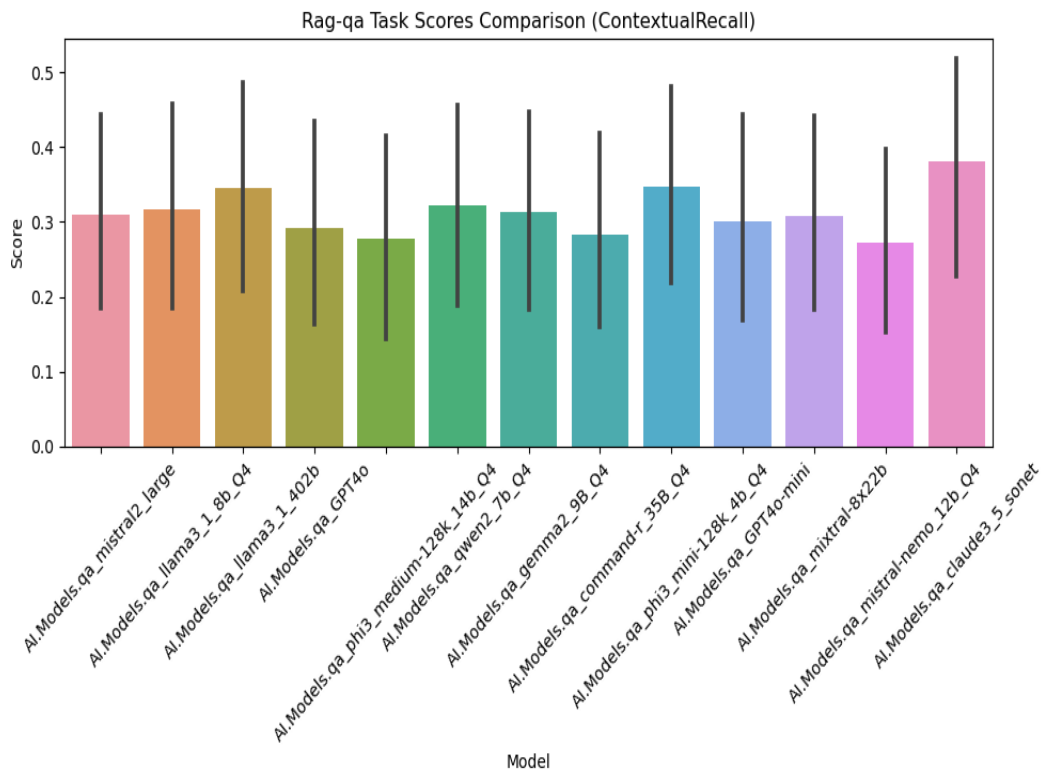
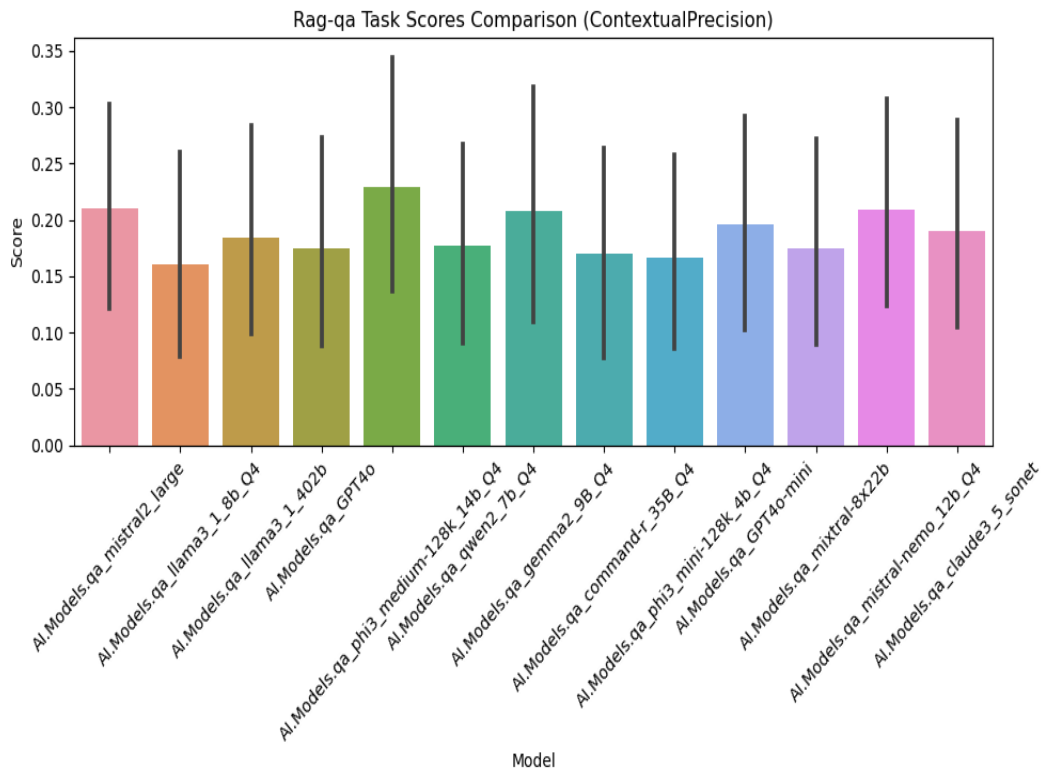


Grouped Results

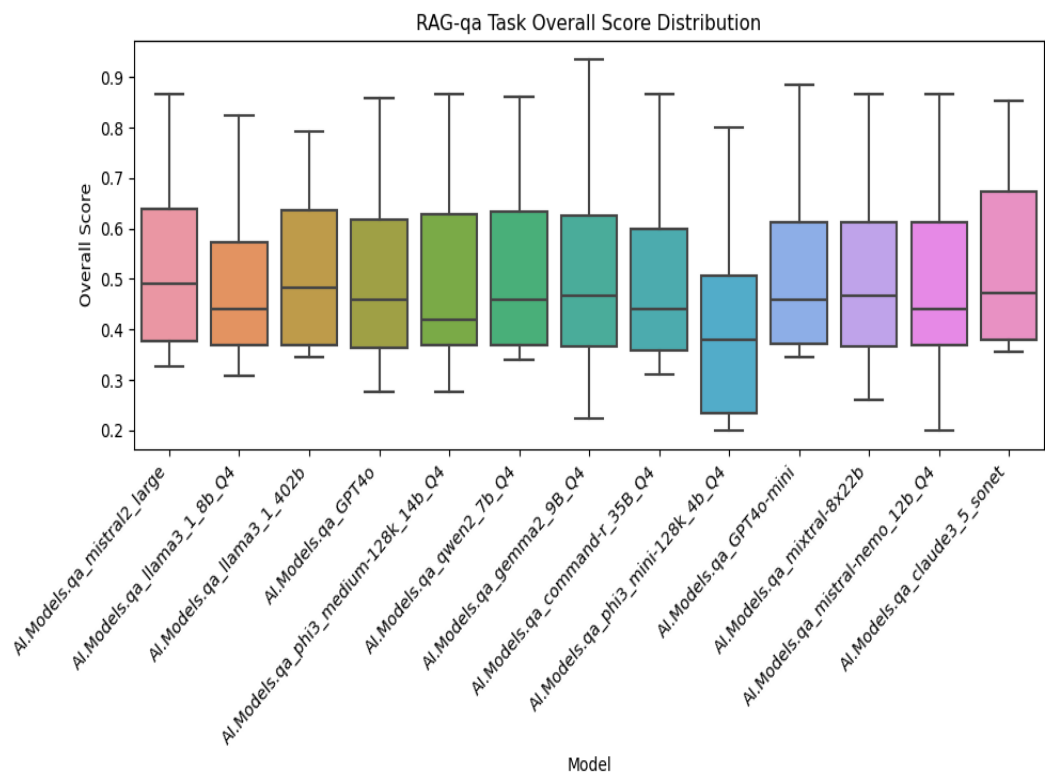


Rag-qa Task Results

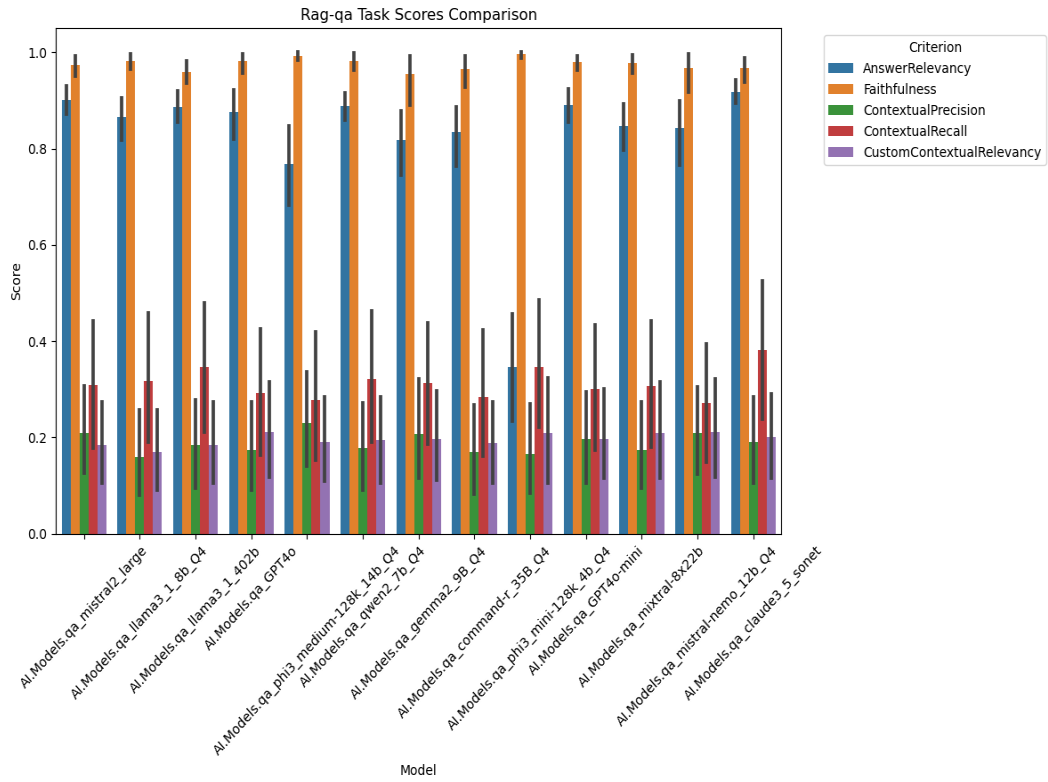




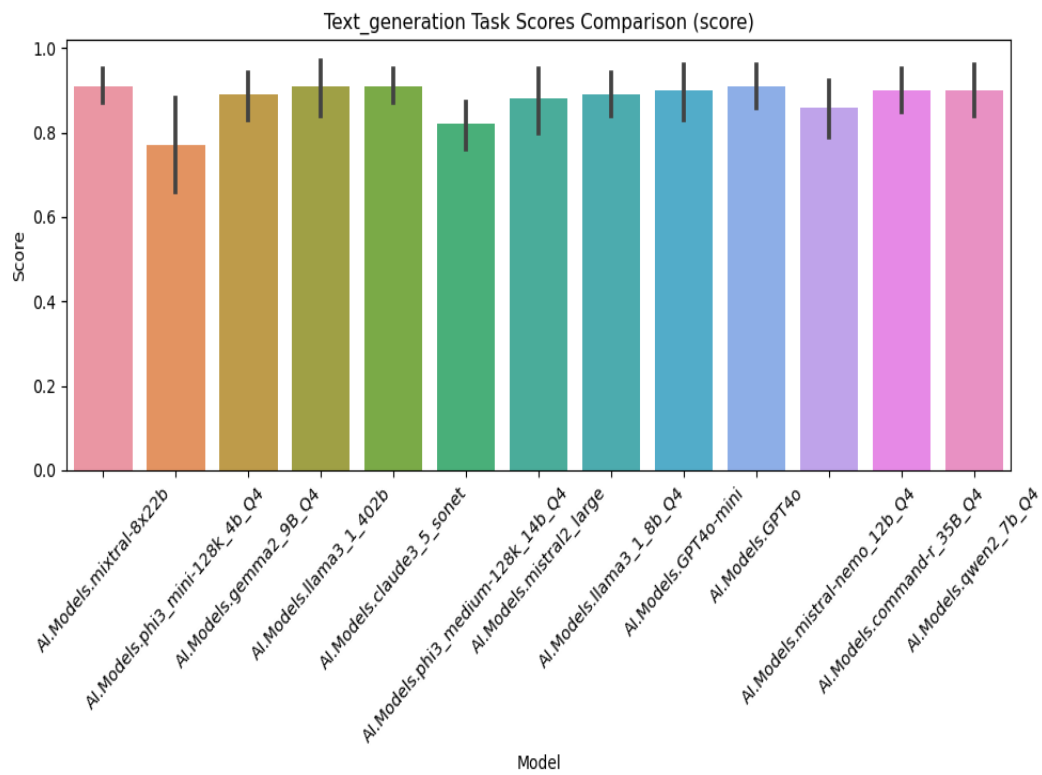
Overall Score Distribution



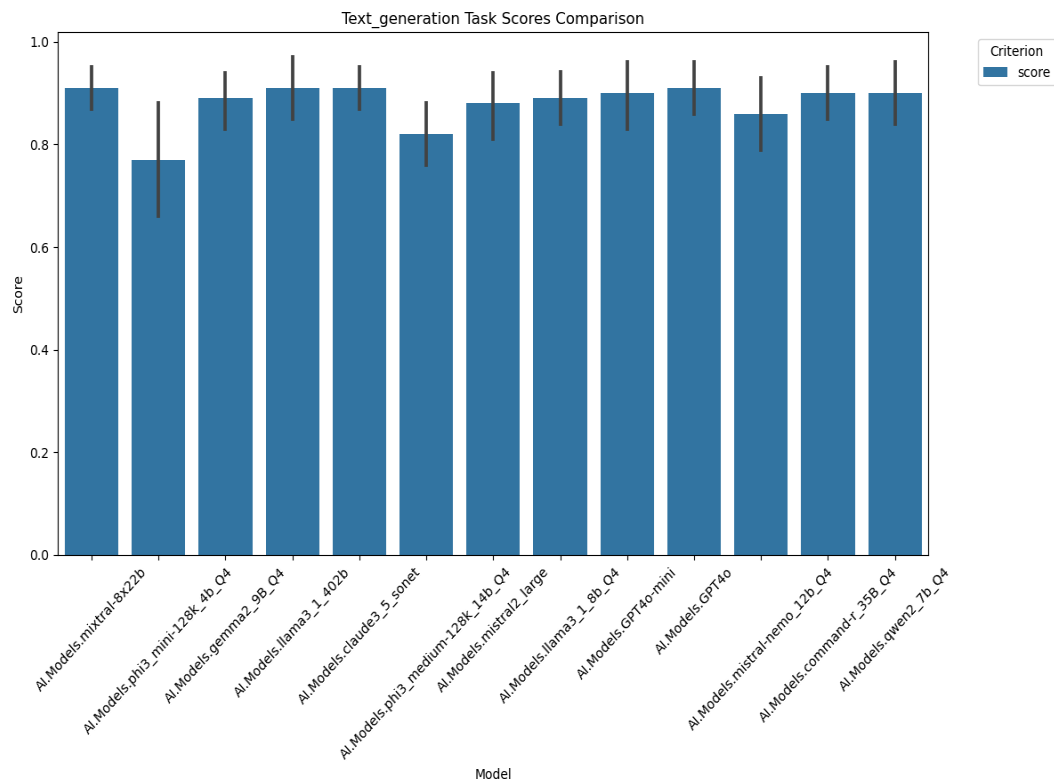
Grouped Results



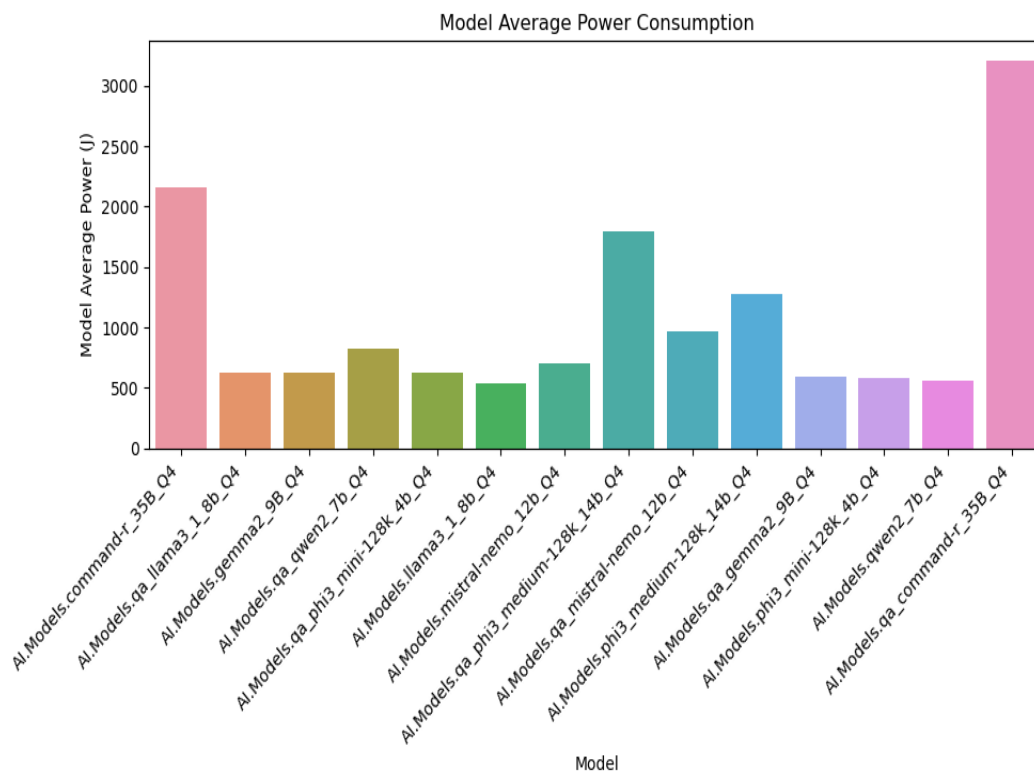
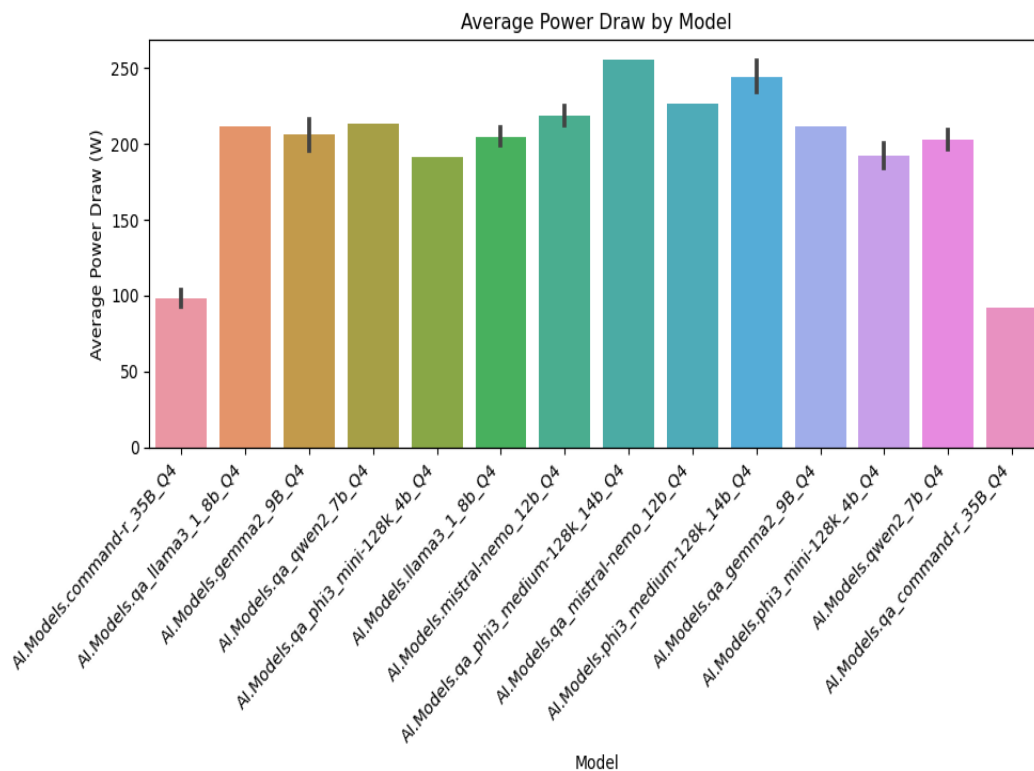
Text_generation Task Results

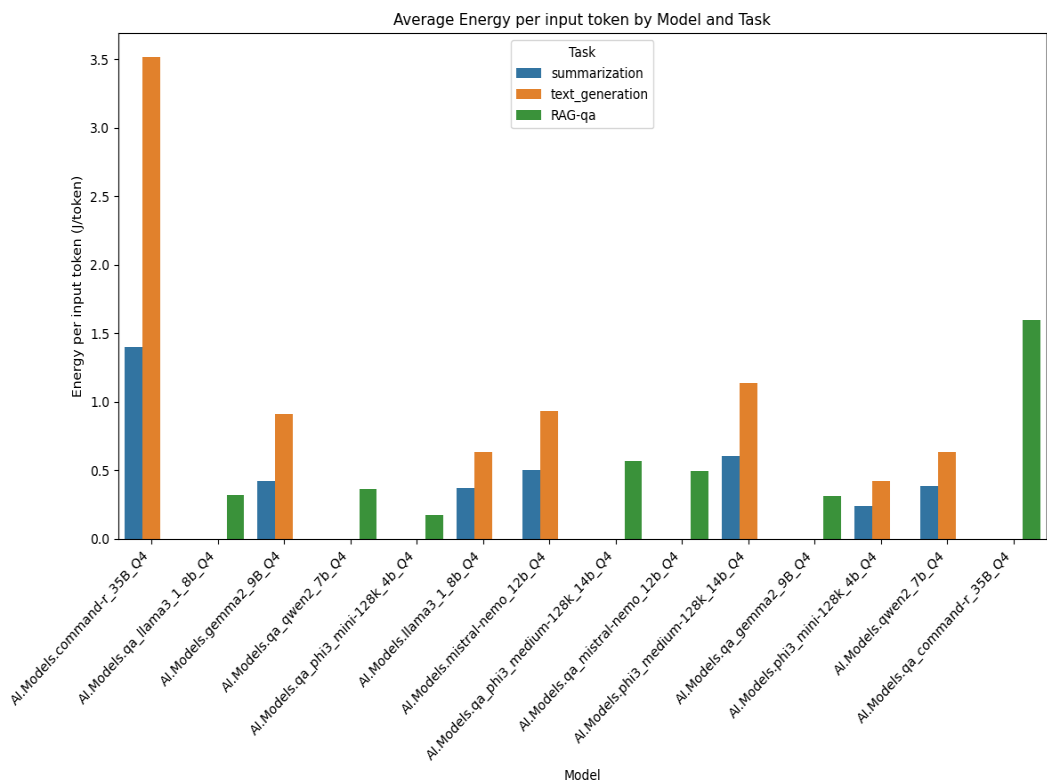
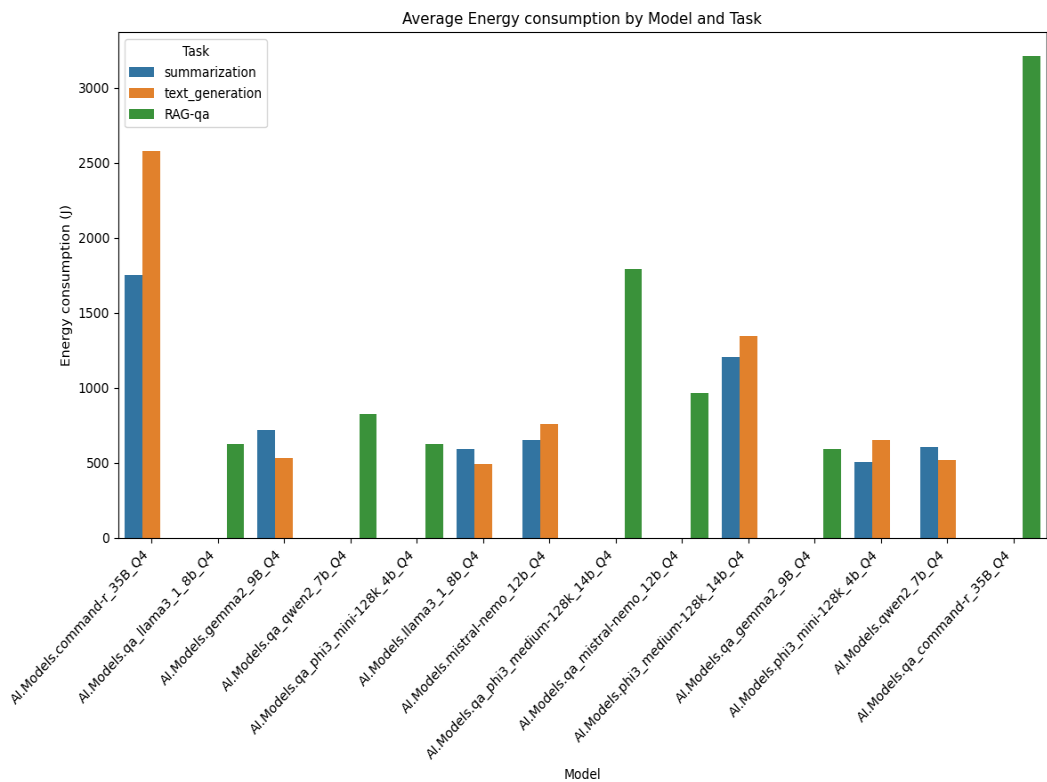


Grouped Results



Power Consumption Results





Evaluation Results

Rag-qa Results

Model: AI.Models.qa_GPT4o

File: qa_001_result.json

Overall_score: 0.8579710144927535

Individual_scores:

AnswerRelevancy: 0.9565217391304348

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant and provides useful information, but it includes a general description of the notification module that does not directly address the specific issue.

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

ContextualPrecision: The score is 0.33 because the relevant node (rank 3) that mentions the 'notifications.enabled' setting in 'xwiki.properties' is ranked lower than irrelevant nodes. For example, the first node discusses notification RSS feed settings but does not address enabling notifications in the 'xwiki.properties' file, and the second node explains overriding notification templates and extension installation without mentioning the 'xwiki.properties' file.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports the expected output. Great job!

CustomContextualRelevancy: The retrieval context explicitly mentions setting 'notifications.enabled' in 'xwiki.properties' to 'true', which matches the expected answer.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.5757142857142857

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.42857142857142855

CustomContextualRelevancy: 0.2

Reasons:

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

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.25 because the relevant node in the fourth rank discusses the 'notifications.enabled' setting in 'xwiki.properties', which is directly relevant to enabling notifications. However, irrelevant nodes in the first, second, third, and fifth ranks discuss topics such as the 'Clear All' link, dependencies for the notifications extension, various ways to receive notifications, and the 'What's

New' application, which should be ranked lower.

ContextualRecall: The score is 0.43 because only parts of the expected output are supported by the nodes in the retrieval context. Specifically, the mention of 'notifications.enabled' in 'xwiki.properties' and receiving notifications via the bell icon are supported, but there are no mentions of subscribing to pages or users, following a user, or the 'network' tab in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties and watching users, but lacks details on subscribing to pages, using the alert menu, and viewing the network tab in the user profile.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.45999999999999996

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and there are no irrelevant statements.

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. For example, the first node discusses how to get a notification RSS feed and general notification settings, but it does not address enabling notifications for one's own actions. The second node lists various notification customization options but does not mention anything about receiving notifications for one's own actions. The third node talks about enabling and disabling the email feature and customizing live email notifications grace time, but it does not address the specific question about receiving notifications for one's own actions. The fourth node provides an overview of different ways to receive notifications in XWiki, but it does not specifically address how to enable notifications for one's own actions. The fifth node discusses overriding default notification templates and installation instructions, which are not relevant to enabling notifications for one's own actions.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about not receiving notifications about one's own activity by default or the ability to disable this filter in the 'Advanced filtering options' section of the notification settings.

CustomContextualRelevancy: The retrieval context mentions notification settings and the ability to enable or disable notifications, but it lacks specific details about disabling the filter for one's own activity and the 'Advanced filtering options' section.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6333333333333333

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3666666666666667

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and there are no irrelevant statements.

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

ContextualPrecision: The score is 0.37 because the relevant nodes in the retrieval context are not ranked higher than the irrelevant nodes. For example, the first node discusses enabling and disabling the notification module and email notifications at the administrator level, which is not directly related to the user's question about disabling notifications for specific pages. Similarly, the second node explains how to get a notification RSS feed and the settings for notifications, but it does not address how to disable notifications for specific pages. These irrelevant nodes should be ranked lower than the third node, which mentions 'Watch and custom filters,' relevant to the user's question. Additionally, the fourth node describes the 'Clear All' link and grouping of notifications, which does not directly address how to disable notifications for specific pages, and should be ranked lower than the fifth node, which mentions 'Choosing the default auto-watched mode,' directly relevant to the user's question.

ContextualRecall: The score is 0.50 because while the retrieval context mentions notifications for pages created or updated by users (supporting sentence 1), it lacks information on 'Automatic page watching' and deleting existing notification filters (unsupporting sentences 3 and 4).

CustomContextualRelevancy: The retrieval context mentions default notification settings and custom filters but lacks details about automatic page watching and changing settings in the user profile.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_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 answer is mostly relevant and provides useful information about configuring OpenID Connect authentication. However, it includes some irrelevant details about the XWiki RESTful API and access rights, which are not directly related to the specific question asked.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to 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 provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed. Similarly, the second node discusses various authentication methods but does not provide the specific property that needs to be set. The third node explains different authentication methods for the XWiki RESTful API, which is not related to the input question. The fourth node is about setting access rights for terminal pages, and the fifth node provides information on logging in and recovering usernames and passwords, neither of which address the configuration needed to automatically enable user accounts upon first login using OpenID Connect.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.
CustomContextualRelevancy: The retrieval context does not mention adjusting the `xwiki.properties` file or the `oidc.enableUser` property.
Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_006_result.json

Overall_score: 0.3546666666666667
Individual_scores:
AnswerRelevancy: 0.84
Faithfulness: 0.9333333333333333
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:
AnswerRelevancy: The score is 0.84 because the answer partially addresses the question but includes multiple irrelevant URLs that do not directly help with customizing user profile fields from the OpenID Connect provider in XWiki.
Faithfulness: The score is 0.93 because the actual output incorrectly states that the `xwiki.properties` file needs to be edited to set the OIDC authenticator class, while the retrieval context specifies that the configuration key xwiki.authentication.authclass must be defined in xwiki.cfg.
ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node provides general information about the OpenID Connect authenticator but does not address customizing user profile fields or mapping additional fields. Similarly, the second node explains different authentication methods for the XWiki RESTful API but does not provide information on customizing or mapping 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 does not contain information about customizing user profile fields or mapping them in the `xwiki.properties` file.
Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_007_result.json

Overall_score: 0.5432051282051283
Individual_scores:
AnswerRelevancy: 0.9583333333333334
Faithfulness: 1.0
ContextualPrecision: 0.25
ContextualRecall: 0.3076923076923077
CustomContextualRelevancy: 0.2
Reasons:
AnswerRelevancy: The score is 0.96 because the answer is highly relevant and provides a clear process for setting up group synchronization between XWiki and an OpenID Connect provider. However, it includes an irrelevant link to the XWiki RESTful API documentation, which is not directly related to the question.
Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.25 because the relevant node providing information on configuring group synchronization (rank 4) is ranked lower than several irrelevant nodes. For example, the first node describes general capabilities and installation details of the OpenID Connect Authenticator, which is not specific to group synchronization. The second node provides links to various configuration guides but lacks detailed steps for group synchronization. Additionally, the third node discusses container authentication, Active Directory, and LDAP, which are unrelated to the input query.

ContextualRecall: The score is 0.31 because while the retrieval context (node 7) mentions configuring group synchronization and properties like `oidc.groups.claim` and `oidc.groups.mapping`, it lacks details on locating and editing the `xwiki.properties` file, configuring the `oidc.groups.separator`, and saving and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions group synchronization and the xwiki.properties file but lacks detailed steps and specific properties like oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.7366233766233766

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.454545454545453

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.93 because the response is mostly relevant and provides useful information about setting attachment restrictions. However, it includes a link to the XWiki syntax guide and discusses using a blacklist to block specific file types, which are not directly relevant to the specific requirement of allowing only images or PDFs up to 10MB.

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 explaining 'how to set the maximum size of an attachment' is ranked first, while all irrelevant nodes discussing 'moving attachments', 'deleting attachments', 'linking to attached files or images', 'linking to external files or images', and 'a table of different types of links' are ranked lower.

ContextualRecall: The score is 0.45 because while the retrieval context covers the steps for setting the maximum attachment size (Step 2, sentences 1-5), it does not mention configuring allowed mimetypes or editing the `xwiki.properties` file (Step 1).

CustomContextualRelevancy: The retrieval context contains steps for configuring maximum attachment size but lacks details on configuring allowed mimetypes.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_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 while the output contains some relevant information, it is diluted by multiple irrelevant statements that do not address the specific question of how to deny script rights to a space administrator.

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. For example, the first node discusses 'methods for checking access rights programmatically' but does not address the specific question. Similarly, the second node explains 'the use of the checkAccess() method' without providing information on denying script rights to a space administrator. The fourth node provides 'details on HTTP methods and status codes for managing attachments and objects in a wiki,' which is unrelated to the question.

ContextualRecall: The score is 0.00 because the expected output sentence cannot be attributed to any nodes in the retrieval context, which focuses on methods for checking access rights without mentioning the inability to deny rights implied by admin rights.

CustomContextualRelevancy: The retrieval context does not mention that rights implied by admin right cannot be denied.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.4682352941176471

Individual_scores:

AnswerRelevancy: 0.9411764705882353

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and provides useful information, but the URL provided does not directly address the configuration of access rights in XWiki, which slightly detracts from the overall relevancy.

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

ContextualPrecision: The score is 0.20 because the relevant node (rank 5) discussing how permissions are set for groups or users is ranked lower than four irrelevant nodes. For instance, the first node discusses the creation of spaces and pages in XWiki without providing information on configuring access rights for teams, which should be ranked lower.

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 mentions setting rights and the concept of spaces and nested pages, but it lacks specific steps for creating groups, setting wiki-wide rights, and inheriting rights for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_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 answer is perfectly relevant and there are no irrelevant statements. 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 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'how to get a notification RSS feed and manage notification settings,' which does not address the issue of enabling notifications in the wiki. Similarly, the second node is about 'overriding default notification templates and installation instructions,' which is also not relevant.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications through `xwiki.properties` or setting `notifications.enabled` to `true`, nor does it provide any information about the `xwiki.properties` file or the `notifications.enabled` setting.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.28888888888888886

Individual_scores:

AnswerRelevancy: 0.4444444444444444

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.44 because multiple URLs provided in the output do not directly address the issue of not receiving notifications, making the response less relevant. However, the score is not lower because the response may still contain some useful information or context related to the topic.

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 discusses 'various applications like Meeting Application, Forum Application, and Ideas Application,' which are not related to notification settings. Similarly, the second node explains 'access rights and API usage,' which is not relevant to enabling notifications or following users/pages.

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 enabling notifications, following pages or users, or configuring 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: 0.275
Individual_scores:
AnswerRelevancy: 0.375
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.38 because the provided URLs and information are not directly related to enabling notifications for one's own activity in XWiki, making much of the output irrelevant to the input query.

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

ContextualPrecision: The score is 0.00 because all the nodes in the retrieval context are irrelevant. The first node discusses 'JWT token generation and its claims', which is unrelated to notifications about user activity. The second node is about 'upgrading wikis and handling invalid extensions', which does not pertain to notifications for user actions. The third node lists 'various XWiki extensions', none of which relate to enabling notifications for a user's own activity. The fourth node is about 'LDAP group mapping and authentication', which is not relevant to the question about enabling notifications for user actions. The fifth node discusses 'form utilization on XWiki.org', which does not address the issue of receiving notifications for one's own activity.

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

CustomContextualRelevancy: The retrieval context does not contain any information related to notification settings or user activity filters.

Question_language: de
Expected_answer_language: de
Answer_language: de

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 answer contains several irrelevant links that do not address the question about disabling notifications. However, the relevant parts of the answer do provide useful information on how to manage notifications.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'XWiki syntax and formatting features,' which is not related

to disabling notifications. Similarly, the second node explains 'JWT Token Authenticator for the LLM AI Application,' which is also unrelated. These irrelevant nodes should be ranked lower.

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 following of pages, notification settings, or deleting notification filters as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.4333333333333333

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.16666666666666666

CustomContextualRelevancy: 0.0

Reasons:

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

Faithfulness: The score is 1.00 because there are no contradictions. Great job staying faithful to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node describes general information about the OpenID Connect Authenticator, its developer, and its installation, but does not address the specific configuration needed to automatically activate user accounts. The second node provides links to various configuration guides and examples but does not mention the specific configuration change needed in `xwiki.properties` to automatically activate user accounts. The third node lists various endpoints and other unrelated configurations, but does not provide the specific configuration needed for automatic user activation. The fourth node discusses container authentication and other unrelated authentication methods, but does not mention the `xwiki.properties` configuration required for automatic user activation. Finally, the fifth node mentions configuration files including `xwiki.properties`, but does not provide the specific property (`oidc.enableUser`) that needs to be set to `true` for automatic user activation.

ContextualRecall: The score is 0.17 because only the mention of 'xwiki.properties' and OpenID Connect authentication in the 2nd node in the retrieval context supports the expected output, while specific instructions, property details, and procedural steps are missing.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.68

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and directly addresses the question without any irrelevant statements. 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 0.20 because the relevant node (rank 5) is ranked much lower than the irrelevant nodes. For example, the first node provides 'prerequisites, installation instructions, and dependencies,' which are not directly related to customizing user profile fields from an OpenID Connect provider. Similarly, the second node describes the 'general functionality of the OpenID Connect Authenticator' without addressing the customization of user profile fields. The third node discusses 'container authentication and other authentication methods,' which are also irrelevant. The fourth node lists 'various endpoints and features such as group synchronization and customization templates,' but does not provide details on mapping additional user profile fields. The relevant node, which mentions 'Configuration files' and 'xwiki.properties,' should be ranked higher.

ContextualRecall: The score is 1.00 because all sentences in the expected output can be directly attributed to the nodes in the retrieval context, particularly the 2nd node.

CustomContextualRelevancy: The retrieval context mentions configuration files and customization but lacks specific details on using 'oidc.user.mapping' in 'xwiki.properties' for mapping user profile fields.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.618937728937729

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9523809523809523

ContextualPrecision: 0.25

ContextualRecall: 0.6923076923076923

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the output is fully relevant and directly addresses the question about setting up group synchronization between XWiki and the OpenID Connect provider.

Faithfulness: The score is 0.95 because the actual output claims there is a specific configuration page for group synchronization, but this is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.25 because the relevant node mentioning 'Group synchronization' and providing a link to a specific section about it is ranked fourth. Irrelevant nodes, such as the first node which only provides general information about OpenID Connect and XWiki, should be ranked lower.

ContextualRecall: The score is 0.69 because while the retrieval context covers most of the key points such as group synchronization, configuring 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', it lacks specific details on opening the 'xwiki.properties' file with a text editor and replacing placeholder values with actual names.

CustomContextualRelevancy: The retrieval context mentions group synchronization and xwiki.properties but lacks detailed steps and configuration examples found in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.8072727272727273

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.6363636363636364

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.80 because the answer provides useful information but includes several irrelevant links and a statement about default settings that do not directly address the input question about configuring attachment restrictions.

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

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node provides information on setting limits on the maximum size of an attachment, which is directly relevant. The subsequent nodes discuss topics like whitelists and blacklists for attachment types (second node), managing document properties (third node), creating new pages (fourth node), and embedding models (fifth node), which are not relevant.

ContextualRecall: The score is 0.64 because while the retrieval context covers the steps to change the maximum upload size in the XWikiPreferences document and mentions repeating the configuration for each (sub)wiki (supporting sentences 6-12), it does not mention configuring allowed MIME types in the `xwiki.properties` file or provide specific details on MIME type restrictions (missing support for sentences 2-5).

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_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 answer partially addresses the question about denying script rights to a space administrator in XWiki, but it includes multiple irrelevant URLs that do not relate to the topic.

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 contains a link to an image and references to various projects, which are not related to the question about denying script rights to a space administrator. Additionally, the second node discusses authentication methods and configurations, including Kerberos and LDAP, which are not relevant to the question about script rights.

ContextualRecall: The score is 0.00 because the expected output discusses XWiki's rights management limitations, which are not covered in any of the nodes in the retrieval context.
CustomContextualRelevancy: The retrieval context does not contain any information about admin rights or their implications in XWiki.
Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_020_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 provided URL does not directly address the specific configuration question multiple times, but the rest of the answer is relevant and helpful.

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 discusses 'various actions that can be performed in XWiki, such as creating children pages and changing page access rights,' but does not provide specific steps for configuring access control as required by the input. Similarly, the second node explains 'the concept of wikis and subwikis, and how to organize content within them,' which does not address the specific access control configuration needed for the input question.

ContextualRecall: The score is 0.00 because none of the sentences in the expected output have a direct match in the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not provide specific steps or details about creating groups, setting wiki-wide and space-level rights, or inheriting rights for child pages as outlined in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_021_result.json

Overall_score: 0.3875
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 0.9375
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 1.00 because the answer is completely relevant and addresses the issue without any irrelevant statements. Great job!

Faithfulness: The score is 0.94 because the actual output incorrectly states that installing extensions offline is not supported, which contradicts the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses getting an RSS feed of notifications and various settings but does not address enabling notifications in the wiki.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about ensuring notifications are activated in the wiki or how to activate them using `xwiki.properties`.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.5266666666666666

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and addresses the issue of not receiving notifications without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.33 because the relevant node, which mentions 'Follow a user' and 'Watched Entities,' is ranked third. Irrelevant nodes, such as the first node discussing 'getting a notification RSS feed and settings for notifications' and the second node about 'overriding default notification templates and installation instructions,' should be ranked lower.

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 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: 0.6142483660130719

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 0.8888888888888888

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.88 because the answer is mostly relevant and addresses the question about enabling notifications for one's own actions, but it includes URLs that are not directly related to the specific query.

Faithfulness: The score is 0.89 because the actual output incorrectly states that you can access the notifications menu by clicking on the bell icon in the top bar, whereas the retrieval context specifies that notifications are received under the Alert entry, represented by a bell, in XWiki 9.3+.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses how to get a notification RSS feed and does not mention anything about enabling notifications for one's own actions. The second node provides a description and links to various notification-related features but does not address the specific issue of enabling notifications for one's own actions. The third node lists various notification preferences macros but does not provide information on how to enable notifications for one's own actions. The fourth node explains how to override default notification templates and install extensions but does not address enabling notifications for one's own actions. The fifth node lists different ways to be notified in XWiki but does not provide specific instructions on enabling notifications for one's own actions.

ContextualRecall: The score is 1.00 because all sentences in the expected output can be directly attributed to the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions default notification settings and filtering options but lacks details about disabling the filter in 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.8516666666666666

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.88 because the main content is relevant and provides useful information on how to disable notifications, but it includes URLs that do not directly address the question.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.58 because the first node discusses getting a notification RSS feed and mentions the notifications macro, which does not directly address how to disable unwanted notifications. Additionally, the fourth node lists various ways to receive notifications but does not provide information on how to disable them. The fifth node discusses overriding default notification templates and installation instructions, which are not relevant to disabling unwanted notifications. However, the second and third nodes are correctly ranked as they mention 'Notifications Auto Watch Preferences,' 'Notifications Custom Filters Preferences,' and 'Default Watch Settings,' which are directly related to modifying notification settings.

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

CustomContextualRelevancy: The retrieval context mentions auto-watched pages and custom notification filters but lacks details on modifying settings and deleting filters.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.3894736842105263

Individual_scores:

AnswerRelevancy: 0.9473684210526315

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 directly, but it includes a link that is not directly related to configuring user account activation via OpenID Connect in XWiki.

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. The first node describes the OpenID Connect Authenticator extension, its installation, and its developer, but it does not provide any information about configuring user activation. The second node discusses container authentication and mentions OpenID Connect Authentication but does not provide details on configuring automatic user activation. The third node provides a general description of the OpenID Connect Authenticator and its synchronization capabilities but does not mention the specific configuration needed for automatic user activation. The fourth node discusses various authentication methods for the XWiki RESTful API but does not address the configuration of automatic user activation with OpenID Connect. The fifth node lists endpoints, group synchronization, customization, and troubleshooting for the OpenID Connect Authenticator but does not mention the configuration for automatic user activation.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the `xwiki.properties` file or setting the `oidc.enableUser` property.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.575

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

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 from an OpenID Connect provider in XWiki. However, it includes URLs to documentation that are not directly related to the specific task, which slightly detracts from the overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'various authentication methods for the XWiki RESTful API, including OIDC,' but does not cover customizing user profile fields. The second node provides 'general information about the OpenID Connect Authenticator for XWiki,' but does not address mapping

additional user profile fields. The third node lists 'configuration properties related to the OpenID Connect provider and endpoints,' without mentioning user profile field mapping. The fourth node explains 'different types of references in XWiki,' which is not relevant to the input. The fifth node discusses 'naming strategies for wiki pages in XWiki,' which is also unrelated.

ContextualRecall: The score is 1.00 because the retrieval context perfectly aligns with all sentences in the expected output, providing comprehensive support for each point mentioned.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6946031746031747

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.42857142857142855

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and provides useful information, but the inclusion of a link about creating new users in XWiki slightly detracts from the focus on configuring group synchronization with an OpenID Connect provider.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect faithfulness!

ContextualPrecision: The score is 0.50 because the second node, which is relevant, is correctly ranked higher and mentions 'default group synchronization' and 'properties in the `xwiki.properties` file, such as `oidc.groups.mapping`'. However, the first, third, fourth, and fifth nodes are irrelevant as they provide 'general information about prerequisites, installation instructions, and dependencies', 'general information about the OpenID Connect authenticator', 'various endpoints, customization options, and troubleshooting steps', and 'a link to more information about managing users and groups in XWiki' respectively, and should be ranked lower.

ContextualRecall: The score is 0.43 because while the retrieval context mentions configuring properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, it lacks detailed instructions such as locating and editing the `xwiki.properties` file, replacing placeholders with actual values, and saving and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions configuring group synchronization in the xwiki.properties file and using properties like oidc.groups.mapping, but lacks detailed steps and examples found in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_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 answer is mostly relevant to the input question, but it includes several irrelevant URLs that do not address the specific query about configuring file upload restrictions.

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 configuring file upload restrictions or size limits. For example, the first node describes 'various macros and their functionalities in XWiki', the second node explains 'how to customize the PDF cover in XWiki', and the third node provides 'instructions on creating a new page via a wanted link'.

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 MIME types or setting maximum upload size for attachments.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.7928571428571429

Individual_scores:

AnswerRelevancy: 0.9642857142857143

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.96 because the answer is highly relevant and addresses the question well, but the inclusion of a link to XWiki syntax documentation slightly detracts from its direct relevance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'default right settler policies,' the second node provides 'information about executing tasks with certain user rights,' the third node discusses 'the Scripting API and checking access rights,' the fourth node lists 'various types of rights and permissions,' and the fifth node explains 'how links are resolved in XWiki.' None of these contexts address the specific question about refusing script rights to an administrator.

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

CustomContextualRelevancy: The retrieval context clearly states that rights implied by an admin right cannot be overridden by a denial at a lower level, matching the expected output that admin rights cannot be refused.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.317948717948718

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 0.6666666666666666

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and provides useful information on configuring permissions in XWiki, but it includes URLs that are not directly related to the specific question asked.

Faithfulness: The score is 0.67 because the actual output repeatedly claims that each team can see and modify spaces, configure permissions, and replicate permissions for each team space. However, the retrieval context clearly states that the notion of space has been removed from the UI and that spaces are only emulated temporarily.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node explains how XWiki handles spaces and pages, but it does not provide information on setting access controls for teams. Similarly, the second node describes various macros and scripting capabilities in XWiki, which are not relevant to configuring access controls for teams.

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

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, defining rights at the wiki or space level, or inheriting rights for child pages.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.5529411764705883

Individual_scores:

AnswerRelevancy: 0.7647058823529411

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.0

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.76 because the answer provides relevant information about authenticating users with access tokens, but it includes multiple irrelevant URLs that do not directly address the question.

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

ContextualPrecision: The score is 0.50 because the first node discusses 'various claims that can be requested at the UserInfo endpoint using an access token,' which is not relevant to authenticating users with access tokens. Additionally, the third node provides 'general information about the Security Module and its ratings,' the fourth node discusses 'cookie protection and IP address usage for login,' and the fifth node discusses 'user profile rights and custom claims for id tokens,' all of which are not relevant to the input. However, the second node provides 'information about configuring authorized applications,' which is relevant to setting up token-based authentication and is correctly ranked.

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.

CustomContextualRelevancy: The retrieval context mentions the 'Authorized Applications' configuration and the use of tokens, but lacks details on JWT claims, Ed25519 key, and specific authenticator setup steps.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_032_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 authenticating users with access tokens. 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. For example, the first node discusses 'group synchronization and customization for OpenID Connect', the second node mentions 'protection levels for cookies and security strategies for failed logins', and the third node explains 'LDAP group mapping and an example of LDAP authentication'. None of these address the use of access tokens for user 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 the LLM-Anwendung (BETA)-Erweiterung, JWT tokens, or the specific authentication steps outlined in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_033_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 answer provides some relevant information but includes URLs that do not directly address the question about authenticating users with access tokens.

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. For example, the first node discusses 'collection resource properties and methods,' which are unrelated to user authentication with access tokens. Similarly, the second node talks about 'various claims at the UserInfo endpoint,' but does not address user authentication with access tokens. These irrelevant nodes should be ranked lower.

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 authenticating users with access tokens, 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_GPT4o-mini

File: qa_001_result.json

Overall_score: 0.8466666666666667

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

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

Faithfulness: The score is 0.90 because the actual output incorrectly specifies `notifications.enabled = true` instead of the correct `notifications.emails.enabled` as mentioned in the retrieval context.

ContextualPrecision: The score is 0.33 because the relevant node mentioning the `notifications.enabled` setting in `xwiki.properties` is ranked third, while irrelevant nodes discussing notification RSS feeds (first node), overriding templates (second node), and notification management features (fourth and fifth nodes) are ranked higher.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports all sentences in the expected output. Great job!

CustomContextualRelevancy: The retrieval context mentions enabling notifications by setting 'notifications.enabled' to 'true' in 'xwiki.properties', which matches the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.5566666666666668

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is completely relevant and there are no irrelevant statements.

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

ContextualPrecision: The score is 0.25 because the relevant node (rank 4) mentions enabling the notification module and email notifications in the `xwiki.properties` file, which is relevant to the expected output. However, irrelevant nodes are ranked higher, such as the first node that discusses how notifications are displayed and managed, the second node that lists dependencies for the notifications extension, and the third node that provides different ways to receive notifications without explaining how to enable them.

ContextualRecall: The score is 0.33 because only a few aspects of the expected output are supported by the nodes in the retrieval context. Specifically, the mention of 'notifications.enabled' in xwiki.properties and the alert menu represented by a bell are covered, but other details such as subscribing to pages or users, following a user, and the 'network' tab in the user profile are not mentioned.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties and watching pages or users, but lacks detailed steps on how to do it and does not cover the network tab or user profile details.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.43999999999999995

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and there are no irrelevant statements in the actual output. Great job!

Faithfulness: The score is 0.90 because the actual output incorrectly states that Notification Settings are accessed under the Alert entry in the top bar, while the retrieval context indicates they are accessed through the notifications menu.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'how to get a notification RSS feed and general notification settings' but does not address enabling notifications for one's own actions. Similarly, the second node lists 'various notification settings and customization options' but does not mention enabling notifications for one's own actions. The third node focuses on 'enabling and disabling email notifications and live email notifications grace time,' which is unrelated to the input query.

ContextualRecall: The score is 0.00 because the retrieval context does not mention that users don't receive notifications about their own activity by default, nor does it mention the ability to disable this filter in the 'Advanced filtering options' section of the user's notification settings.

CustomContextualRelevancy: The retrieval context mentions notification settings and default behaviors but does not explicitly cover disabling the filter in 'Advanced filtering options' or targeting specific users.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6133333333333334

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3666666666666667

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Reasons:

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

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining accuracy and consistency!

ContextualPrecision: The score is 0.37 because irrelevant nodes are ranked higher than relevant ones. For instance, the first node discusses enabling/disabling the notification module and email notifications at the platform level, which is not directly relevant to the user's question about disabling notifications for specific pages. Similarly, the second node explains how to get an RSS feed of notifications and mentions the settings menu, but it does not provide specific information about disabling notifications for certain pages. These should be ranked lower than the third node, which mentions 'Watch and custom filters' and provides a link to more details, and the fifth node, which mentions 'Choosing the default auto-watched mode,' both of which are directly relevant to the user's question.

ContextualRecall: The score is 0.50 because while the retrieval context nodes mention default notification settings related to user actions on pages they created or updated (supporting sentences 1 and 2), there is no information about changing 'Automatic page watching' settings or deleting notification filters in the 'Custom Filters' list (unsupporting sentences 3 and 4).

CustomContextualRelevancy: The retrieval context mentions 'auto-watched mode' and 'custom filters' but lacks details on automatic page watching, notification settings, and deleting filters.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.3464285714285714

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 0.8571428571428571

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the core information about configuring user accounts with OpenID Connect is mostly relevant, but the inclusion of unrelated links slightly detracts from the overall relevance.

Faithfulness: The score is 0.86 because the actual output incorrectly states that the `xwiki.authentication.createuser=empty` property in the `xwiki.cfg` file is related to OpenID Connect, while the retrieval context specifies that this configuration is for Container Authentication.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed to enable user accounts upon first login. Similarly, the second node discusses container authentication without providing details on configuring the `xwiki.properties` file or enabling user accounts upon first login.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the xwiki.properties file or the oidc.enableUser property.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_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 provides useful information related to XWiki and OpenID Connect, but includes links and details that are not directly relevant to customizing user profile fields.

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. For example, the first node discusses the OpenID Connect Authenticator extension and its general features but does not provide information on customizing user profile fields or mapping additional fields. Similarly, the second node explains various authentication methods for the XWiki RESTful API, including OIDC, but does not address customizing user profile fields or mapping additional fields.

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

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.5120735785953177

Individual_scores:

AnswerRelevancy: 0.9565217391304348

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.15384615384615385

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.96 because the answer is highly relevant and provides clear guidance on configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes a minor irrelevant mention of the XWiki RESTful API, which is not directly related 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 0.25 because the relevant node (rank 4) addressing group synchronization is ranked lower than several irrelevant nodes. For example, the first node discusses

'general features' of the OpenID Connect Authenticator, the second node provides 'links to various configuration guides', and the third node talks about 'different authentication methods', none of which are specific to configuring group synchronization.

ContextualRecall: The score is 0.15 because the retrieval context only partially supports the expected output, specifically in configuring group synchronization and defining mappings, while missing many other critical details such as locating and editing the `xwiki.properties` file, setting various properties, and providing examples.

CustomContextualRelevancy: The retrieval context mentions group synchronization and the xwiki.properties file but lacks detailed steps and specific property configurations found in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.85

Individual_scores:

AnswerRelevancy: 0.95

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.95 because the answer is highly relevant and provides useful information, but the inclusion of a link to XWiki syntax documentation slightly detracts from the focus on setting upload restrictions.

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

ContextualPrecision: The score is 1.00 because the relevant node that provides detailed steps on setting the maximum size of an attachment in the XWikiPreferences document is ranked first, while all irrelevant nodes discussing topics like moving attachments, whitelists and blacklists, linking to external files, and linking to attachments and anchors are ranked lower.

ContextualRecall: The score is 0.50 because while the retrieval context covers the steps for configuring the maximum attachment size (Step 2), it does not mention the steps for configuring allowed mimetypes (Step 1).

CustomContextualRelevancy: The retrieval context contains steps for setting maximum attachment size and mentions mimetype restrictions, but lacks specific details on configuring the `xwiki.properties` file and the exact mimetype allow list.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.384

Individual_scores:

AnswerRelevancy: 0.92

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and provides useful information on denying script rights to a space administrator. However, the inclusion of URLs that are not directly related to the question slightly detracts from its 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'methods for checking access rights in code', which does not address the specific question. Similarly, the second node talks about 'the usage of the checkAccess() method', and the third node provides 'examples of access checks using hasAccess() and checkAccess() methods', none of which are relevant to denying script rights to a space administrator. Additionally, the fourth and fifth nodes describe 'HTTP methods and status codes for managing attachments and objects in a wiki', which are also not pertinent to the input question.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about denying rights that are implied by admin rights or the inability to deny such rights, making the expected output unsupported.

CustomContextualRelevancy: The retrieval context does not mention that rights implied by admin right cannot be denied.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.46400000000000001

Individual_scores:

AnswerRelevancy: 0.92

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and provides useful information, but it includes URLs that are not directly related to configuring view and edit permissions in XWiki, 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 0.20 because the relevant node discussing 'how permissions are set for groups and users' is ranked fifth, while irrelevant nodes such as 'how pages and spaces are created and managed' (rank 1), 'parent/child relationship and nested pages' (rank 2), 'various macros available' (rank 3), and 'specification for XWiki link syntax' (rank 4) are ranked higher.

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.

CustomContextualRelevancy: The retrieval context mentions setting rights and creating spaces but lacks specific steps on creating groups for each team and setting rights at different levels.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_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 input, but the term 'Benutzereinstellungen' is too vague and does not provide specific information relevant to the issue mentioned.

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. For example, the first node discusses 'getting a notification RSS feed and settings for notifications,' which does not address the issue of the notification bell not being available.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications through the `xwiki.properties` file or setting `notifications.enabled` to `true`.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_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 while the response addresses the issue of not receiving notifications, it includes multiple irrelevant URLs that do not directly pertain to 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'various applications like Meeting Application, Forum Application, and Ideas Application,' which are not related to enabling notifications. The second node talks about 'access rights and executing tasks with specific user rights,' which is also unrelated to receiving 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 enabling notifications, following pages or users, or configuring notifications in xwiki.properties.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.3473684210526316
Individual_scores:
AnswerRelevancy: 0.7368421052631579
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.74 because the output contains several URLs that are not relevant to enabling notifications for one's own actions, which detracts from the overall relevancy. However, the score is not lower because the output likely contains some useful information or correct URLs related to the query.

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. For example, the first node is about 'token generation and JWT tokens', which is unrelated to notifications for user activity. Similarly, the second node discusses 'database migration and invalid extensions', which does not pertain to notifications for user activity.

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

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

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_014_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 answer provides relevant information on how to disable notifications, but it includes multiple irrelevant URLs that do not address the question about disabling notifications.

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. For example, the first node discusses 'XWiki Syntax and formatting features,' which does not address disabling notifications. Similarly, the second node describes 'an extension for authenticating users with JSON Web Tokens,' which is also unrelated.

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 mentioned in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_015_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 answer is completely relevant and directly addresses the question 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node describes general information about the OpenID Connect Authenticator but does not provide any information on how to automatically activate user accounts upon first login. Similarly, the second node provides a table of contents and links to various sections related to the OpenID Connect Authenticator but does not address the specific configuration needed to automatically activate user accounts.

ContextualRecall: The score is 0.20 because only the mention of 'xwiki.properties' and 'OpenID Connect' in the 2nd node in the retrieval context aligns with the expected output. However, there are no specific instructions or details about updating properties, uncommenting, saving the file, restarting the XWiki instance, or ensuring new user accounts are automatically activated.

CustomContextualRelevancy: The retrieval context does not contain information about modifying the 'xwiki.properties' file or the 'oidc.enableUser' property.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.68

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

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

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

ContextualPrecision: The score is 0.20 because the relevant node, which mentions configuration files including 'xwiki.properties', is ranked fifth. The first four nodes are irrelevant as they primarily provide links to prerequisites, describe general capabilities, mention various authentication methods, and list endpoints without detailing how to map additional user profile fields.

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

CustomContextualRelevancy: The retrieval context mentions the xwiki.properties file but lacks details on customizing user profile fields using oidc.user.mapping.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.5685714285714286

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.6428571428571429

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the output is fully 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 that the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node provides general information about OpenID Connect and its integration with XWiki but does not provide specific steps for configuring group synchronization. The second node mentions allowing XWiki to authenticate with an OpenID Connect provider but does not mention group synchronization or the specific configuration steps required. The third node describes container authentication and other authentication methods but does not address group synchronization with OpenID Connect. The fourth node lists various configuration guides and files but does not provide the detailed steps for setting up group synchronization between XWiki and an OpenID Connect provider. The fifth node provides links to configuration guides for specific systems and mentions the `xwiki.properties` file but does not detail the steps for configuring group synchronization.

ContextualRecall: The score is 0.64 because while the retrieval context nodes mention key configuration elements like 'xwiki.properties', 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping', they lack specific instructions such as opening the file with a text editor, replacing placeholders, and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions group synchronization but lacks detailed steps and properties configuration found in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.884705882352941

Individual_scores:

AnswerRelevancy: 0.8235294117647058

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.82 because the main content addresses the input question, but the inclusion of irrelevant URLs lowers the score.

Faithfulness: The score is 1.00 because there are no contradictions. Great job staying faithful to the retrieval context!

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node provides 'detailed steps on setting limits on the maximum size of an attachment,' which is directly relevant to the input query. Irrelevant nodes, such as the second node discussing 'MIME type whitelists and blacklists,' the third node about 'managing documents,' the fourth node on 'creating new pages in XWiki,' and the fifth node on 'properties of document collections,' are all correctly ranked lower.

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

CustomContextualRelevancy: The retrieval context includes steps to change the maximum size of an attachment and mentions MIME type restrictions, but lacks specific details about configuring `xwiki.properties` and the exact MIME types allowed.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_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 answer addresses the question about denying script rights to a space administrator, but several irrelevant links about XWiki skins, authentication, LDAP authentication, and forms reduce 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 0.00 because the nodes in the retrieval context are irrelevant to the input. For example, the first node is a list of image links and references, which does not address the question about denying script rights to a space administrator. The second node is about configuring Kerberos and LDAP authentication in XWiki, which is unrelated to the question about script rights for a space administrator. The third node discusses mail templates for password reset and username recovery, which is not relevant to the question about script rights.

ContextualRecall: The score is 0.00 because the sentence 'Leider wird dies in XWiki nicht unterstützt, Rechte, die durch das Admin-Recht impliziert sind, können nicht verweigert werden.' is not found in any part of the retrieval context, which primarily contains references to images, LDAP configuration, and form usage.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_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 provides useful information about configuring access rights in XWiki, but it includes several irrelevant links to general documentation pages that do not directly address the specific question about team access configuration.

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 configuring access control for teams in XWiki. For example, the first node discusses 'various actions that can be performed in XWiki,' the second node provides 'general information about wikis, panels, and their configurations,' and the third node lists 'various macros and their functionalities in XWiki.' These 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 node(s) in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting wiki-wide rights, or managing access rights for different teams as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_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 completely relevant and addresses the issue about the notification bell not being available despite the extension being installed.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses getting a notification RSS feed and does not mention enabling notifications in the wiki settings. Similarly, the second node explains how to override default notification templates and install extensions, which is not related to enabling notifications in the wiki settings.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about ensuring notifications are activated in the wiki or how to activate them, nor does it provide any information about defining the property `notifications.enabled` in `xwiki.properties` to `true`.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.6316666666666666

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.625

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and addresses the issue of not receiving notifications without any irrelevant information. Great job!

Faithfulness: The score is 0.90 because the actual output incorrectly states that you need to disable a filter to not receive notifications for actions done by yourself, whereas the retrieval context specifies that by default, you won't receive such notifications without mentioning any filter.

ContextualPrecision: The score is 0.33 because the relevant node, which mentions 'Watched Entities' and 'Follow a user' (rank 3), is ranked lower than several irrelevant nodes. For example, the first node discusses getting a notification RSS feed and settings for notifications, which is not directly related to enabling notifications or following pages/users.

ContextualRecall: The score is 0.62 because while the retrieval context covers following pages, users, and using the alert menu (nodes 3), it lacks details on enabling notifications through the `xwiki.properties` file and viewing followed users in the user profile network tab.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting `notifications.enabled` to `true` in `xwiki.properties`, and the steps to follow pages or users are not fully covered.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.6927777777777778

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 0.875

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.89 because the answer is mostly relevant and provides useful information on enabling notifications, but it includes URLs that are not directly related to the question about notifications for one's own actions.

Faithfulness: The score is 0.88 because the actual output incorrectly states that users need to click on the bells to receive alerts for their own actions, whereas the retrieval context clarifies that notifications are received by default when users are mentioned or when actions are performed on pages they created or updated.

ContextualPrecision: The score is 0.50 because the relevant node ranked 2nd, which mentions 'Filter your own events,' is correctly identified as relevant. However, irrelevant nodes ranked 1st, 3rd, 4th, and 5th should be ranked lower as they do not address notifications for one's own activity. For example, the first node discusses getting a notification RSS feed and general notification settings but does not address notifications for one's own activity.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports the expected output, with nodes in the retrieval context clearly explaining the notification settings and filtering

options.

CustomContextualRelevancy: The retrieval context mentions default notification settings and the ability to filter notifications, but it lacks details on disabling the filter in 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7013725490196079

Individual_scores:

AnswerRelevancy: 0.8235294117647058

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.75

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.82 because the provided URLs are not directly relevant to the question about how to disable notifications. However, the answer still contains useful information related 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 0.33 because the relevant node in the third rank mentions 'Auto Watch' and 'Default Watch Settings,' which are pertinent to modifying auto-watch settings and disabling notifications. However, irrelevant nodes in the first, second, fourth, and fifth ranks discuss topics like getting a notification RSS feed, notification preferences macros, different notification methods, and overriding default notification templates, which do not address how to disable notifications or modify auto-watch settings.

ContextualRecall: The score is 0.75 because the retrieval context partially matches the expected output. Specifically, nodes in the retrieval context mention 'Auto Watch', 'Default Watch Settings', 'Auto Watch Preferences', and 'Custom Filters Preferences', which relate to the automatic watching of pages, modifying notification settings, and managing custom filters. However, the context does not explicitly mention notifications about changes to your work.

CustomContextualRelevancy: The retrieval context mentions auto watch preferences and custom filters but lacks details on modifying settings and removing filters.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_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 answer is highly relevant and addresses the question about configuring OpenID Connect in XWiki. However, it includes a link to the XWiki RESTful API documentation, which is not directly relevant to the specific configuration issue.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context provide general information about the OpenID Connect Authenticator, container authentication, or various authentication methods, but none of them address the specific configuration needed to automatically activate user accounts. For example, the first node provides general information about the OpenID Connect Authenticator, such as its developer, active installs, and license, but does not address the specific configuration needed to automatically activate user accounts.

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 adjusting the `xwiki.properties` file or setting the `oidc.enableUser` property to true.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.5599999999999999

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.80 because the answer provides useful information but includes several irrelevant URLs that do not directly address the customization of user profile fields using OpenID Connect in XWiki.

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. For example, the first node discusses various authentication methods for the XWiki RESTful API, including OIDC, but does not address how to map user profile fields from an OpenID Connect provider to XWiki user properties. The second node provides information about the OpenID Connect Authenticator extension for XWiki, but it does not explain how to map user profile fields from an OpenID Connect provider to XWiki user properties. The third node lists various properties related to OpenID Connect configuration in the `xwiki.properties` file, but it does not mention the `oidc.user.mapping` property or how to map user profile fields.

ContextualRecall: The score is 1.00 because all the information in the expected output is accurately supported by the 3rd node in the retrieval context, covering all necessary details without any discrepancies.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6128205128205129

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.23076923076923078

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is highly relevant and provides detailed information on configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes a link to creating new users, 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 0.50 because the relevant node that explains how to enable group synchronization by adding the claim `xwiki_groups` to the property `oidc.userinfoclaims` in the `xwiki.properties` file is ranked second. However, irrelevant nodes that provide general information about prerequisites, installation instructions, dependencies, developer details, active installs, license, endpoints, customization options, troubleshooting steps, and managing users and groups are ranked first, third, fourth, and fifth.

ContextualRecall: The score is 0.23 because while the 2nd node in the retrieval context mentions configuring properties like ``oidc.groups.claim``, ``oidc.groups.separator``, and ``oidc.groups.mapping``, it lacks details on locating and editing the ``xwiki.properties`` file, providing examples, and saving changes.

CustomContextualRelevancy: The retrieval context mentions configuring group synchronization in `xwiki.properties` and using `oidc.groups.mapping`, but lacks detailed steps and examples found in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_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 main content is relevant and addresses the question about restricting file uploads, but the inclusion of unrelated URLs slightly detracts from the overall relevancy.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses 'various macros and features related to notifications, office document viewing, PDF table of contents, and other unrelated functionalities,' which does not address file upload restrictions or attachment size limits. The second node focuses on 'customizing the PDF cover by overriding a template and adding a logo,' which is also unrelated. The third node explains 'how to create a new page by adding a wanted link and using templates,' which does not provide information on configuring file uploads. The fourth node lists 'various macros and their functionalities,' which does not address the input query. Lastly, the fifth node provides 'detailed information on image styling parameters,' which is unrelated to file upload restrictions or attachment size limits.

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 MIME types or setting maximum upload size for attachments.
Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_029_result.json

Overall_score: 0.7302392344497608
Individual_scores:
AnswerRelevancy: 0.8421052631578947
Faithfulness: 0.9090909090909091
ContextualPrecision: 0
ContextualRecall: 1.0
CustomContextualRelevancy: 0.9
Reasons:

AnswerRelevancy: The score is 0.84 because the answer is mostly relevant but includes URLs that do not directly address the question about refusing script rights to a space administrator.
Faithfulness: The score is 0.91 because the actual output claims that you can deny the script right for a group, but the retrieval context does not mention the script right or its management.
ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'default right settler policies' which does not address the specific question about denying script rights to an administrator.
ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the information provided in the retrieval context, specifically the 1st node.
CustomContextualRelevancy: The retrieval context confirms that rights implied by an admin right cannot be overridden, aligning closely with the expected answer about admin rights not being deniable.
Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_030_result.json

Overall_score: 0.515
Individual_scores:
AnswerRelevancy: 0.875
Faithfulness: 1.0
ContextualPrecision: 0.7
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.88 because the provided answer is mostly relevant and addresses the configuration of spaces and permissions in XWiki. However, the inclusion of a URL that does not directly address the specific configuration needs slightly detracts from the overall relevance.
Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!
ContextualPrecision: The score is 0.70 because the first node is relevant and provides information on managing spaces in XWiki. However, the second node, which discusses macros, is irrelevant and should be ranked lower. Additionally, the third node, which talks about parent/child relationships without specific access control details, and the fourth node, which explains link specifications, are also

irrelevant and should be ranked lower. The fifth node is relevant and should be ranked 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.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.46000000000000001

Individual_scores:

AnswerRelevancy: 0.6

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.60 because while the answer may contain some relevant information about authenticating users with access tokens, it is significantly diluted by multiple irrelevant URL statements that do not directly address the question.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect faithfulness!

ContextualPrecision: The score is 0.50 because the second node, which is relevant as it discusses 'Authorized Applications Configuration' for token-based authentication, is ranked higher than some irrelevant nodes. However, the first node, which discusses profile claims and access tokens without providing information about authenticating users with access tokens, is ranked higher than the relevant node.

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

CustomContextualRelevancy: The retrieval context mentions 'Authorized Applications' and some configuration details but lacks specifics on JWT claims, Ed25519 key, and token usage.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_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 was perfectly relevant and addressed the question directly. Great job!

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node is about 'group synchronization and customization for OpenID Connect,' which is not related to authenticating users with access tokens. Similarly, the second node discusses 'protection levels for cookies and security strategies for failed login attempts,' which are also irrelevant.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication, JWT, or the specific steps outlined in the expected answer.

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 answer partially addresses the question about authenticating users with access tokens, but includes irrelevant URLs that do not directly answer the query.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'properties and methods related to a collection resource in an application,' which is not relevant to user authentication with access tokens. Similarly, the second node talks about 'various claims that can be requested by an access token,' but does not address user authentication. These irrelevant nodes should be ranked lower.

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 authenticating users with access tokens, JWT, or the specific steps and requirements mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_claude3_5_sonet

File: qa_001_result.json

Overall_score: 0.8533333333333333

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9333333333333333

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and addresses the question directly without any irrelevant information.

Faithfulness: The score is 0.93 because the actual output incorrectly refers to a setting `notifications.enabled = true` in the `xwiki.properties` file, while the retrieval context only mentions `notifications.emails.enabled`.

ContextualPrecision: The score is 0.33 because the relevant node, which discusses enabling notifications by setting `notifications.enabled` to true in the `xwiki.properties` file, is ranked third. Irrelevant nodes, such as the first node discussing notification RSS feeds and settings, and the second node explaining how to override default notification templates, are ranked higher.

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: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.5225

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9375

ContextualPrecision: 0.25

ContextualRecall: 0.125

CustomContextualRelevancy: 0.3

Reasons:

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

Faithfulness: The score is 0.94 because the actual output claims the setting for enabling notifications should be `notifications.enabled = true`, but the retrieval context does not mention this setting.

ContextualPrecision: The score is 0.25 because the relevant node in retrieval context (rank 4) that mentions the 'notifications.enabled' setting in 'xwiki.properties' is ranked lower than irrelevant nodes. For example, the first node mainly describes the 'Clear All' link and how notifications are grouped, which is not relevant to enabling notifications or subscribing to pages or users.

ContextualRecall: The score is 0.12 because only the mention of 'notifications.enabled' in xwiki.properties was found in the nodes in the retrieval context, while other details about enabling notifications, subscribing to pages or users, and following users were not present.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties and the alert menu but lacks details on subscribing to pages or users, and the user profile network tab.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.39555555555555555

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 0.8888888888888888

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.89 because the answer is mostly relevant and provides useful information, but it includes some irrelevant details about standard CMS functionality and why notifications for one's own actions are not sent, which do not directly address the specific question of enabling such notifications in XWiki.

Faithfulness: The score is 0.89 because the actual output incorrectly claims that XWiki does not send notifications to users regarding their own activities by default, while the retrieval context specifies that users receive notifications when mentioned or when actions are performed on pages they created or updated.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context discuss various aspects of notifications such as RSS feeds, customization options, email features, and notification templates, but none address how to enable notifications for one's own actions.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about not receiving notifications about one's own activity or the ability to disable this filter in the 'Advanced filtering options' section of the notification settings in the user profile.

CustomContextualRelevancy: The retrieval context mentions notification settings and filters but does not explicitly state that users don't receive notifications about their own activity by default or how to disable this filter in advanced settings.

Question_language: en

Expected_answer_language: en

Answer_language: fr

File: qa_004_result.json

Overall_score: 0.6025641025641025

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8461538461538461

ContextualPrecision: 0.3666666666666667

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and there are no irrelevant statements in the actual output. Great job!

Faithfulness: The score is 0.85 because the actual output incorrectly claims that notifications are enabled by default for various types of events, while the retrieval context specifies that notifications are only enabled by default for mentions and actions on pages created or updated by the user. Additionally, the actual output incorrectly states that turning on the 'Own event filter' reduces notifications, whereas the retrieval context indicates that turning off the filter allows users to receive notifications for their own actions.

ContextualPrecision: The score is 0.37 because the relevant nodes are not ranked higher than the irrelevant nodes. For example, the first node discusses enabling and disabling the notification module and email notifications but does not address how to disable notifications for specific pages or user settings. Similarly, the second node talks about getting an RSS feed of notifications and customizing notification settings, but it does not provide information on disabling notifications for specific pages or user settings. These irrelevant nodes should be ranked lower than the third node, which mentions 'Watch and custom filters' and 'Enabling notifications for all users by default,' and the fifth node, which includes 'Choosing the default auto-watched mode' and 'Create global notification filters,' both directly relevant to managing and disabling notifications for specific pages.

ContextualRecall: The score is 0.50 because only some parts of the expected output are supported by the nodes in the retrieval context. Specifically, the retrieval context mentions notifications related to pages users created or updated (2nd node) and accessing notification settings (2nd node), but it does not mention notifications about changes on your work or deleting existing notification filters in the 'Custom Filters' list.

CustomContextualRelevancy: The retrieval context mentions settings and filters but lacks details on automatic page watching and major modifications.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.3646153846153847

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 0.9230769230769231

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.90 because the majority of the response is relevant to configuring OpenID Connect authentication in XWiki. However, it includes some irrelevant links about the XWiki RESTful API and access rights, which do not directly address the specific configuration issue.

Faithfulness: The score is 0.92 because the actual output incorrectly claims to modify the properties `xwiki.authentication.createuser=1` and `xwiki.authentication.trustAuthentication=1`, while the retrieval context only mentions `xwiki.authentication.createuser=empty` for Container Authentication.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node 'does not provide any information on configuring user account enablement', the second node 'does not provide relevant configuration details for enabling user accounts', and the third node 'does not provide specific configuration instructions for enabling user accounts'.

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

CustomContextualRelevancy: The retrieval context does not contain information about adjusting the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: en

Expected_answer_language: en

Answer_language: fr

File: qa_006_result.json

Overall_score: 0.38666666666666666

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the response is mostly relevant and addresses the question about customizing user profile fields from an OpenID Connect provider in XWiki. However, it

includes links to the XWiki RESTful API documentation and the Administration Application extension, which are not directly related to the specific query.

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. For example, the first node describes 'general features and installation details of the OpenID Connect Authenticator' but does not provide information on customizing user profile fields or mapping additional fields. Similarly, the second node discusses 'various authentication methods for the XWiki RESTful API' without mentioning customization of user profile fields. All nodes fail to address the specific query about mapping additional 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 does not contain any information about customizing user profile fields or mapping them in the xwiki.properties file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.5127692307692308

Individual_scores:

AnswerRelevancy: 0.96

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.15384615384615385

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.96 because the answer is highly relevant and detailed, but it includes a link to the XWiki RESTful API documentation, which is not directly relevant to 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 in the fourth position provides information about group synchronization, including the property `oidc.groups.mapping` and the need to add the claim `xwiki_groups` to the property `oidc.userinfoclaims` in the `xwiki.properties` file. However, irrelevant nodes in the first, second, third, fifth, and sixth positions, which discuss topics like the OpenID Connect Authenticator extension, various configuration guides, different authentication methods, customization options, and authentication methods for the XWiki RESTful API, are ranked higher than the relevant node.

ContextualRecall: The score is 0.15 because the retrieval context only partially supports the expected output, specifically mentioning the `oidc.groups.claim` and `oidc.groups.mapping` properties, but lacks comprehensive details on configuring group synchronization, locating and editing the `xwiki.properties` file, and other specific configurations.

CustomContextualRelevancy: The retrieval context mentions group synchronization and properties like `oidc.groups.mapping` but lacks detailed steps and examples found in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.8029090909090909

Individual_scores:

AnswerRelevancy: 0.96

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.454545454545453

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.96 because the answer is highly relevant and provides a clear solution, but it includes a minor irrelevant statement about XWiki syntax that is not directly related to configuring attachment settings.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node provides detailed steps on how to set the maximum size of an attachment, which is directly relevant to configuring the maximum attachment size to 10MB. Irrelevant nodes, such as those discussing moving attachments and refactoring pages (second node), whitelists and blacklists for attachment downloads (third node), linking to external files and backlinks indexation (fourth node), and linking to attachments and anchors (fifth node), are correctly ranked lower.

ContextualRecall: The score is 0.45 because while the retrieval context provides detailed steps for configuring the maximum attachment size (Steps 2.1 to 2.5), it does not mention configuring allowed mimetypes or editing the `xwiki.properties` file (Step 1).

CustomContextualRelevancy: The retrieval context contains steps to set the maximum attachment size and mentions mimetype restrictions but lacks specific details on configuring the `xwiki.properties` file and the exact mimetype allowList.

Question_language: en

Expected_answer_language: en

Answer_language: fr

File: qa_009_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 answer is mostly relevant and provides useful information on denying script rights to a space administrator. However, it includes URLs that are not directly related to the question, which slightly detracts from the overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node describes the checkAccess and hasAccess methods, which are related to access control but do not address the specific question. Similarly, the second node discusses the usage of the checkAccess() method and its importance in security checkpoints, but it does not provide information about denying script rights to a space administrator.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention that rights implied by admin right cannot be denied.

CustomContextualRelevancy: The retrieval context does not mention that rights implied by admin right cannot be denied, which is the key information in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.44504201680672273

Individual_scores:

AnswerRelevancy: 0.9428571428571428

Faithfulness: 0.8823529411764706

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and provides a clear solution to configuring access rights in XWiki, but it includes a URL that does not directly address the specific question.

Faithfulness: The score is 0.88 because the actual output incorrectly suggests that edit rights can be denied at the wiki level and only allowed at the space level, whereas the retrieval context clarifies that denying edit rights at the wiki level prevents page creation unless explicitly allowed at the Page & Children level.

ContextualPrecision: The score is 0.20 because the relevant node (rank 5) discusses 'explicit setting of permissions for groups or users and the implications of the EDIT right', which is directly related to configuring access rights for teams. However, irrelevant nodes ranked higher (ranks 1-4) discuss topics such as 'how pages and spaces are created and managed', 'parent/child relationship and Nested Pages', 'various macros and their purposes', and 'link syntax', none of which address access control or rights management.

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.

CustomContextualRelevancy: The retrieval context mentions setting rights and the concept of spaces and pages, but it lacks specific steps about creating groups, setting wiki-wide rights, and inheriting rights for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_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 addresses the issue 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 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses getting an RSS feed of notifications, which does not address enabling notifications in the wiki settings. The second node is about overriding default notification templates and installation instructions, which is also not relevant. The third node lists various settings and features related to notifications but does not mention enabling notifications. The fourth node discusses customizing notification emails and creating global notification filters, which is unrelated. Finally, the fifth node explains the 'Clear All' link and grouping notifications, but does not mention enabling notifications in the wiki settings.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications via the `xwiki.properties` file or setting `notifications.enabled` to `true`, nor does it provide any information about the `xwiki.properties` file or the specific setting `notifications.enabled`.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.354545454545455

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 contains several irrelevant links that do not address the issue of not receiving notifications, but it still provides some relevant information.

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 'various XWiki applications like Meeting Application, Forum Application, and Ideas Application,' which does not relate to enabling notifications. Similarly, the second node describes 'access rights and the use of the AuthorExecutor role,' which is also unrelated to 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.

CustomContextualRelevancy: The retrieval context does not contain any information about enabling notifications, following pages or users, or configuring notification settings in xwiki.properties.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_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 answer provides some relevant information but includes multiple URLs that are not directly related to enabling notifications for one's own activities, which detracts from 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node is about 'token generation and JWT claims', which is unrelated to notifications for user activity. Similarly, the second node discusses 'database migration and invalid extensions', which does not pertain to notifications for user activity.

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

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_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 provided URLs are not relevant to the question about disabling notifications, which detracts from the overall relevance of the answer.

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

ContextualPrecision: The score is 0.00 because all the nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'XWiki Syntax and formatting features', which does not help in answering the question about disabling notifications. Similarly, the second node describes 'an extension for authenticating users with JSON Web Tokens', which is also unrelated.

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

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.64

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and directly addresses the question about configuring XWiki for automatic user activation with 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 0.20 because the relevant node (rank 5) is ranked lower than four irrelevant nodes. The first node provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed. The second node includes links to prerequisites and installation instructions but does not mention the specific configuration required. The third node mentions various configuration guides for specific systems but does not provide the specific configuration needed. The fourth node discusses container authentication and Active Directory integration but does not mention the specific configuration needed.

ContextualRecall: The score is 1.00 because the retrieval context perfectly matches the expected output, specifically mentioning the 'xwiki.properties' file and its configuration for OpenID Connect.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.6733333333333333

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.16666666666666666

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and addresses the question 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 0.17 because the relevant node (rank 6) is ranked lower than several irrelevant nodes. For example, the first node provides general information about prerequisites and installation instructions, which does not address customization of user profile fields. Similarly, the second node describes general functionality without mentioning user profile field customization. These irrelevant nodes should be ranked lower than the relevant node that mentions configuration files and the 'oidc.user.mapping' property.

ContextualRecall: The score is 1.00 because all sentences in the expected output are perfectly supported by the nodes in the retrieval context, showing a complete and accurate match.

CustomContextualRelevancy: The retrieval context mentions xwiki.properties and customization but lacks detailed steps on mapping user profile fields and restarting XWiki.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.62

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.1

Reasons:

AnswerRelevancy: The score is 1.00 because the response is fully relevant and directly addresses the question about setting up group synchronization between XWiki and 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node provides general information about prerequisites, installation instructions, and dependencies for the OpenID Connect Authenticator but does not address the specific steps for configuring group synchronization. Similarly, the second node gives a brief description of the OpenID Connect Authenticator and its general features but does not provide any specific instructions on setting up group synchronization.

ContextualRecall: The score is 1.00 because every sentence in the expected output is well-supported by the nodes in the retrieval context, specifically the 2nd node which comprehensively covers group synchronization, configuration properties, and the 'xwiki.properties' file.

CustomContextualRelevancy: The retrieval context mentions group synchronization but lacks detailed steps and specific 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: 0.6673376623376623

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 0.625

ContextualPrecision: 1.0

ContextualRecall: 0.45454545454545453

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.86 because the answer is mostly relevant and addresses the question about limiting uploads to images or PDFs with a maximum size of 10 MB, but it includes irrelevant URLs that do not directly contribute to the solution.

Faithfulness: The score is 0.62 because the actual output incorrectly claims restrictions on file types and sizes that are not mentioned in the retrieval context, such as a 10 MB limit and specific MIME types like application/pdf.

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes, ensuring precise and accurate retrieval.

ContextualRecall: The score is 0.45 because while the retrieval context provides detailed steps for configuring the maximum attachment size (sentences 7-11), it lacks information on configuring allowed MIME types (sentences 1-6).

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

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_019_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 answer is mostly relevant but includes multiple irrelevant URLs that do not address the question about denying script rights to a space administrator.

Faithfulness: The score is 1.00 because there are no contradictions. 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 'a list of image links and references', the second node discusses 'Kerberos authentication and LDAP configuration', and the third node describes 'mail templates for password reset and username retrieval'. These topics do not address the question about denying script rights to a space administrator in XWiki.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about XWiki not supporting certain rights or the inability to refuse rights implied by admin rights.

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

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_020_result.json

Overall_score: 0.3741935483870968
Individual_scores:
AnswerRelevancy: 0.8709677419354839
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.87 because the answer is mostly relevant to the question about configuring XWiki permissions for different teams. However, it includes several links to general XWiki documentation that do not directly address the specific configuration question, which slightly reduces the score.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'various actions that can be performed in XWiki' but does not provide specific steps on configuring access control for teams. The second node describes 'the concept of wikis and subwikis' without addressing the specific access control requirements for teams. The third node lists 'various macros and their functionalities' without providing information related to

configuring access control for teams. The fourth node mentions 'applications and extensions' without guidance on setting up access control for teams. The fifth node discusses 'language preferences and translations,' which is unrelated to configuring access control for teams.

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

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting wiki-wide rights, setting area-level rights, or inheriting rights for child pages.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_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 addresses the issue directly without any irrelevant statements. 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'how to get a notification RSS feed and settings related to notifications,' which does not address enabling notifications in the wiki.

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.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications by 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: 0.7184848484848485

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9090909090909091

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.75

CustomContextualRelevancy: 0.6

Reasons:

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

Faithfulness: The score is 0.91 because the actual output incorrectly claims 'Vous ne recevez pas de notifications.' while the retrieval context states that users do receive notifications by default when they are mentioned or when actions are performed on pages they created or updated.

ContextualPrecision: The score is 0.33 because the relevant node, which mentions 'Watched Entities' and 'Follow a user,' is ranked third. Irrelevant nodes, such as the first node discussing RSS feed settings and the second node about overriding notification templates, are ranked higher.

ContextualRecall: The score is 0.75 because while the retrieval context covers following pages, users, and using the alert menu, it lacks details on activating notifications via the `notifications.enabled` property in `xwiki.properties`.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.7688888888888889

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and addresses the question about enabling notifications for one's own actions, but it includes an unnecessary link to XWiki syntax documentation, which is not directly related to the query.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.50 because the relevant node ranked 2nd mentions 'Filter your own events,' which is directly relevant to the question about enabling notifications for one's own actions. However, irrelevant nodes ranked 1st, 3rd, 4th, and 5th discuss topics such as getting a notification RSS feed, various notification preferences macros, overriding default notification templates, and various ways to be notified in XWiki, which are unrelated to enabling notifications for one's own actions.

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

CustomContextualRelevancy: The retrieval context mentions default notification settings and filtering options, but lacks details about disabling the filter in 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.6806666666666666

Individual_scores:

AnswerRelevancy: 0.92

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.75

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and provides useful information on how to disable notifications. However, it includes URLs that do not directly address the question about disabling notifications, 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 relevant node (rank 3) is ranked lower than multiple irrelevant nodes. For example, the first node discusses getting a notification RSS feed and various notification settings, but it does not address how to disable automatic notifications or modify notification settings in the user's profile. Similarly, the second node lists various notification preferences macros but does not explain how to disable automatic notifications or modify notification settings in the user's profile.

ContextualRecall: The score is 0.75 because the retrieval context partially supports the expected output. Specifically, the 1st node in the retrieval context supports sentences 1 and 3 of the expected output by mentioning default notification settings and the notifications menu. The 2nd node supports sentence 4 by referencing notification preferences and custom filters. However, the retrieval context does not specifically mention being notified about changes to one's own work, which affects the completeness of the match.

CustomContextualRelevancy: The retrieval context mentions auto watch and custom filters preferences but lacks details on modifying the auto-watch setting and removing filters.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.471304347826087

Individual_scores:

AnswerRelevancy: 0.9565217391304348

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant and addresses the main question about configuring OpenID Connect in XWiki. However, it includes a minor irrelevant reference to the XWiki RESTful API, which slightly detracts from the overall focus.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For instance, the first node provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed to automatically activate user accounts. Similarly, the second node discusses container authentication without detailing the required configuration for automatic user account activation. The third node describes the functionality of the OpenID Connect Authenticator but does not mention the specific property `oidc.enableUser` or the `xwiki.properties` file. The fourth node discusses various authentication methods for the XWiki RESTful API but lacks information on configuring automatic user account activation with OpenID Connect. Lastly, the fifth node lists various endpoints and customization options but does not provide the specific configuration needed to automatically activate user accounts.

ContextualRecall: The score is 0.40 because while there is some alignment with the general concept of OpenID Connect and automatic user authentication in the retrieval context, it lacks specific details about the property `oidc.enableUser`, its value, and the steps to save the file and restart the XWiki instance.

CustomContextualRelevancy: The retrieval context does not mention adjusting the `xwiki.properties` file or setting the `oidc.enableUser` property.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.6405797101449275

Individual_scores:

AnswerRelevancy: 0.8695652173913043

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.87 because the answer is mostly relevant to the question about customizing user profile fields from an OpenID Connect provider in XWiki. However, it includes several irrelevant links to documentation that do not directly address the specific customization query.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect faithfulness!

ContextualPrecision: The score is 0.33 because the relevant node (third node) is ranked lower than irrelevant nodes (first, second, fourth, and fifth nodes). For example, the first node discusses 'various authentication methods for the XWiki RESTful API' which is not related to mapping user profile fields, and the second node provides information about the 'OpenID Connect Authenticator extension' but does not address mapping user profile fields.

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

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.7904761904761906

Individual_scores:

AnswerRelevancy: 0.9523809523809523

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.95 because the main response is highly relevant and provides useful information about configuring group synchronization between XWiki and an OpenID Connect provider. However, the inclusion of a link to the user guide for creating new users slightly detracts from the focus on group synchronization.

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 second node, which directly addresses group synchronization by mentioning the `xwiki.properties` file and the `oidc.groups.claim` property, is ranked

lower than the first node. The first node provides general information about prerequisites, installation instructions, and dependencies for the OpenID Connect Authenticator, which is less relevant to the specific steps for configuring group synchronization.

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

CustomContextualRelevancy: The retrieval context mentions configuring group synchronization in xwiki.properties and using oidc.groups.mapping, but lacks detailed steps and examples found in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_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 provided URL does not directly address the input question about configuring file upload restrictions. However, the rest of the response is relevant and helpful.

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 retrieval context are irrelevant to the input. For example, the first node 'describes various macros and functionalities in XWiki but does not address file upload restrictions or MIME type configurations', and the second node 'explains how to customize the PDF cover in XWiki, which is unrelated to setting file upload restrictions or MIME types'.

ContextualRecall: The score is 0.00 because none of the steps or configurations mentioned in the expected output can be attributed to the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain 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: 0.77

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 provides useful information but includes URLs that do not directly address the specific question about revoking script rights from a space administrator.

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 question about refusing script rights to an administrator. For example, the first node discusses 'default right settler additional policies,' the second node includes 'technical details about access rights and execution contexts,' and the third node provides 'sample code and API details,' none of which address the specific issue.

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

CustomContextualRelevancy: The retrieval context clearly states that rights implied by the administrator cannot be overridden, matching the expected answer.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.38461538461538464

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and provides useful information on configuring access rights in XWiki. However, it includes some irrelevant URL links that do not directly address 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. The first node, ranked 1, describes technical details of XWiki spaces and pages without addressing access rights for teams. The second node, ranked 2, discusses macros and scripting languages, which are not related to configuring access rights. The third node, ranked 3, explains the parent/child relationship and nested pages, but does not cover access rights configuration. The fourth node, ranked 4, focuses on the transition to the new nested pages model, without providing guidance on setting access rights for teams.

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

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, defining rights at the wiki or space level, or inheriting rights for child pages.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.7036363636363637

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.82 because the main content addresses the question about authenticating users with access tokens effectively, but the presence of multiple irrelevant URLs prevents the score from being higher.

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 first node discusses 'access to various user claims at the UserInfo endpoint and OAuth 2.0 refresh tokens,' which is not directly related to authenticating users with access tokens and should be ranked lower. Additionally, the third node provides 'general information about permissions and security modules,' which is also not directly related and should be ranked lower. However, the second node provides relevant information about 'configuring Authorized Applications,' which is pertinent to the input query and is correctly ranked higher.

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

CustomContextualRelevancy: The retrieval context mentions 'Authorized Applications' and some configuration details, but lacks specific information about JWT claims, token generation, and the exact authenticator class.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_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 answer is fully relevant and directly addresses the question about authenticating users with access tokens. Great job!

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For instance, the first node discusses 'group synchronization and customization for OpenID Connect,' which is unrelated to authenticating users with access tokens. Similarly, the second node is about 'protection levels for cookies and security strategies for failed logins,' which does not relate to the input query.

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

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication, JWT tokens, or the specific steps to enable and configure the LLM application extension.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_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 question about authenticating users with access tokens, but it includes several unrelated links that do not directly address the query.

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. For example, the first node describes properties and methods related to collections in a wiki, which is not relevant to authenticating users with access tokens. The second node discusses access tokens and claims at the UserInfo endpoint, but does not provide information on authenticating users with access tokens in the context of the application LLM.

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 authenticating users with access tokens, JWT, or the specific steps and requirements 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: 0.8666666666666668
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 0.3333333333333333
ContextualRecall: 1.0
CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and there are no irrelevant statements.

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

ContextualPrecision: The score is 0.33 because the relevant node (third node) discussing the `notifications.enabled` setting in the `xwiki.properties` file is ranked lower than irrelevant nodes. For example, the first node talks about getting a notification RSS feed and various settings related to notifications, which is not relevant to enabling notifications in `xwiki.properties`. Similarly, the second node discusses overriding default notification templates and installation instructions for extensions, which is also not relevant.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports the expected output. Great job!

CustomContextualRelevancy: The retrieval context explicitly mentions the 'notifications.enabled' setting in 'xwiki.properties' to enable notifications, matching the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.54

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.25

CustomContextualRelevancy: 0.2

Reasons:

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

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

ContextualPrecision: The score is 0.25 because the relevant node, which mentions the 'notifications.enabled' setting in 'xwiki.properties', is ranked fourth. Irrelevant nodes, such as the first node discussing visual aspects of notifications and the second node listing dependencies for the notifications extension, are ranked higher.

ContextualRecall: The score is 0.25 because only a few elements from the expected output, such as 'notifications.enabled' and watching a page or its children, are mentioned in the nodes in the retrieval context, while most other details are missing.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties and the alert menu but lacks details on subscribing to pages or users and viewing the network tab.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.4492307692307692

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.85 because the answer is mostly relevant and provides useful information, but it includes URLs that do not directly address the question about enabling notifications for own actions in XWiki.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'getting a notification RSS feed and general notification settings' but does not address enabling notifications for one's own actions. Similarly, the second node lists 'various notification settings and customization options' but does not mention how to enable notifications for

one's own actions.

ContextualRecall: The score is 0.00 because the retrieval context does not mention that users do not receive notifications about their own activity by default (sentence 1) and does not mention the ability to disable this filter in the 'Advanced filtering options' section of the notification settings (sentence 2).

CustomContextualRelevancy: The retrieval context mentions default notification settings and the ability to enable/disable notifications, but it lacks specific details about disabling the filter for one's own activity and the 'Advanced filtering options' section.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5816666666666668

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.875

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Reasons:

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

Faithfulness: The score is 0.88 because the actual output incorrectly states that email notifications can be disabled through the User Profile page, while the retrieval context specifies that it must be done by editing the configuration variable in xwiki.properties and restarting the server.

ContextualPrecision: The score is 0.33 because the relevant node (rank 3) is ranked lower than multiple irrelevant nodes. For example, the first node is about 'enabling or disabling the notification module and email notifications,' which is not directly related to disabling notifications for specific pages. Similarly, the second node discusses 'getting a notification RSS feed and the settings for notifications,' which does not provide information on disabling notifications for specific pages.

ContextualRecall: The score is 0.50 because while some aspects like automatic addition to the watch list (sentence 1) and accessing notification settings (sentence 3) are supported by nodes in the retrieval context, other details such as being notified about changes (sentence 2) and deleting existing notification filters (sentence 4) are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions auto-watched mode and notification settings but lacks details on modifying pages, automatic addition to watch list, and deleting custom filters.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.3931034482758621

Individual_scores:

AnswerRelevancy: 0.9655172413793104

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.97 because the answer is highly relevant and provides useful information for configuring OpenID Connect in XWiki. However, it includes a minor irrelevant statement about the XWiki RESTful API, 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. For example, the first node 'describes the OpenID Connect Authenticator extension and its general information, but it does not provide any configuration details related to enabling user accounts upon first login.'

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 adjusting the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.3096153846153846

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 0.625

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and provides useful information on customizing user profile fields from the OpenID Connect provider in XWiki. However, it includes links to the XWiki RESTful API documentation and the Administration Application extension, which are not directly relevant to the specific question asked.

Faithfulness: The score is 0.62 because the actual output incorrectly references the `xwiki-oidc-authentication.xml` file, whereas the retrieval context only mentions configurations through xwiki.cfg and xwiki.properties.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node describes the OpenID Connect authenticator but lacks information on customizing user profile fields or mapping additional fields. The second node explains different authentication methods for the XWiki RESTful API but does not mention customizing user profile fields or mapping additional fields from OpenID Connect. The third node discusses in-wiki configuration for the OIDC authenticator but does not provide details on mapping additional user profile fields or customizing user profile fields. The fourth node is about managing users and groups in XWiki, including creating, deleting, and editing user profiles, but it does not address mapping additional user profile fields from OpenID Connect. The fifth node mentions various authentication methods, including OpenID Connect, but does not provide information on customizing user profile fields or mapping additional fields from OpenID Connect.

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 customizing user profile fields or mapping them in the xwiki.properties file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.5097628458498024

Individual_scores:

AnswerRelevancy: 0.8260869565217391

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.2727272727272727

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.83 because the answer provides relevant information about configuring group synchronization between XWiki and an OpenID Connect provider, but it includes several irrelevant statements about the XWiki RESTful API, user profile editing, and features unrelated to the synchronization process.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.25 because the relevant node providing specific information on group synchronization (rank 4) is ranked lower than several irrelevant nodes. For example, the first node describes the OpenID Connect Authenticator extension but lacks details on configuring group synchronization, and the second node lists various configuration guides without detailing group synchronization steps.

ContextualRecall: The score is 0.27 because while the retrieval context mentions configuring group synchronization and the `oidc.groups.claim` and `oidc.groups.mapping` properties, it lacks details on locating and editing the `xwiki.properties` file, setting the `oidc.groups.separator` property, and saving and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions group synchronization and the xwiki.properties file but lacks detailed instructions on configuring oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping properties.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.7666666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.83 because the provided URLs are not directly relevant to the input, which specifically asks about ensuring file type and size limits for uploads. However, the rest of the information is pertinent and useful.

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 are ranked higher than irrelevant nodes, ensuring the most pertinent information is prioritized. Great job!

ContextualRecall: The score is 0.50 because while the retrieval context covers the steps for setting the maximum attachment size (Steps 2.1 to 2.5), it does not mention configuring allowed mimetypes or editing the `xwiki.properties` file (Step 1).

CustomContextualRelevancy: The retrieval context includes steps to change the maximum attachment size and mentions mimetype restrictions, but does not provide specific details on configuring the allowList for mimetypes.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.3571428571428571

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.79 because the answer provides some relevant steps to deny script rights to a space administrator, but it includes irrelevant URLs and incorrectly references Slack, which detracts from its overall accuracy and relevance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect faithfulness!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node describes methods for checking access rights programmatically but does not address the specific question of denying script rights to a space administrator. Similarly, the second node discusses the usage of the checkAccess() method and its importance for security but does not mention the denial of script rights to a space administrator.

ContextualRecall: The score is 0.00 because the retrieval context does not mention or explicitly state that certain rights cannot be denied if implied by admin rights.

CustomContextualRelevancy: The retrieval context does not mention that rights implied by admin right cannot be denied.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.3964705882352941

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 0.7

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.88 because the provided URL does not directly address the configuration of access rights in XWiki, which is the main focus of the input. However, the rest of the information is relevant and helpful.

Faithfulness: The score is 0.70 because the actual output incorrectly states that the notion of space has been removed from the UI, preventing setting permissions and creating new spaces, which contradicts the retrieval context.

ContextualPrecision: The score is 0.20 because the relevant node (rank 5) about 'setting permissions and configuring access rights in XWiki' is ranked lower than the irrelevant nodes. For example, the first node discusses 'how users create pages and spaces in XWiki,' which is not related to configuring view or edit rights, and should be ranked lower.

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 mentions setting rights and creating spaces but lacks specific steps for creating groups, setting wiki-wide rights, and inheriting rights for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_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 contains some relevant information but also includes general affirmations and unrelated details about WhatsApp Web functionality, which do not directly address the specific issue of the notification bell not being available.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses getting a notification RSS feed and settings for notifications, but it does not address enabling notifications in the wiki or the specific setting `notifications.enabled`.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications through the `xwiki.properties` file or setting `notifications.enabled` to `true`, which are key points in the expected output.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_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 issue of not receiving notifications was addressed, but multiple irrelevant links were included, which detracted from the overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'various XWiki applications and extensions,' which is not related to enabling notifications or following users/pages. Similarly, the second node talks about 'access rights and executing tasks with specific user rights,' which also does not pertain to the input.

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

CustomContextualRelevancy: The retrieval context does not contain any information about enabling notifications, following pages or users, or managing notifications in user profiles.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_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 actual output contains several irrelevant URLs that do not address the input question about enabling notifications for one's own actions. However, the score is not lower because the relevant information is still present and partially addresses the user's query.

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. For example, the first node is about 'token generation and JWT claims', which is unrelated to notifications for user activity. The second node discusses 'database migration and invalid extensions', which does not pertain to user notification settings. Additionally, the third node lists 'various extensions and their functionalities', none of which are related to enabling notifications for user actions.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context discuss notifications, user activity, or mention 'Erweiterte Filteroptionen' or user notification settings.

CustomContextualRelevancy: The retrieval context does not contain any information about notification settings or user activity filters.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_result.json

Overall_score: 0.34444444444444444

Individual_scores:

AnswerRelevancy: 0.7222222222222222

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.72 because while the answer addresses the question about deactivating notifications, it includes multiple irrelevant URLs that do not pertain to the query.

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 the nodes in retrieval context ranked from 1 to 5 discuss 'XWiki Syntax, verbatim content, and quotations', 'an extension for authenticating users with JSON Web Tokens', 'various management and API resources for the LLM Application', 'information for developers on building applications with XWiki', and 'database migration and invalid extensions during an upgrade', which are all unrelated to disabling notifications.

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

CustomContextualRelevancy: The retrieval context does not contain any information related to automatic page tracking, 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: 0.44000000000000006

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 answer is completely relevant and addresses the question directly without any irrelevant statements. 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 0.00 because all nodes in the retrieval context are irrelevant. The first node provides general information about the OpenID Connect Authenticator, but does not address the specific configuration needed. The second node mentions the ability of XWiki to authenticate using OpenID Connect, but lacks the specific configuration details. The third node discusses container authentication and mentions the extension, but does not provide the needed configuration details. The fourth node provides links to various guides and mentions different authentication methods, but lacks the specific configuration details. The fifth node provides a table of contents and links to sections related to the OpenID Connect Authenticator, but does not provide the needed configuration details.

ContextualRecall: The score is 0.20 because only the mention of 'xwiki.properties' in the 2nd node in the retrieval context supports the expected output, while the rest of the instructions and details are not supported by the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about modifying the 'xwiki.properties' file or setting the 'oidc.enableUser' property.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.62

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.1

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and directly addresses the question without any irrelevant statements. 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node provides general information about the OpenID Connect Authenticator, including prerequisites, dependencies, and installation instructions, but does not address the specific question about customizing user profile fields. The second node describes the general functionality of the OpenID Connect Authenticator and provides some metadata about the extension, but it does not provide information on how to customize user profile fields. The third node discusses container authentication and mentions the OpenID Connect Authenticator, but it does not provide details on customizing user profile fields. The fourth node lists various endpoints and customization options for the OpenID Connect Authenticator, but it does not specifically address how to map additional user profile fields. The fifth node provides a table of contents and links to various sections related to the OpenID Connect Authenticator, but it does not directly answer the question about customizing user profile fields.

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

CustomContextualRelevancy: The retrieval context mentions 'xwiki.properties' but lacks details on 'oidc.user.mapping' and specific configuration steps.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.64

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

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

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node provides general information about OpenID Connect authentication without specific steps for group synchronization. The second node describes the OpenID Connect Authenticator extension but lacks details on configuring group synchronization. The third node discusses other authentication methods like Active Directory and LDAP, not group synchronization with OpenID Connect. The fourth node mentions group synchronization but does not provide detailed steps or settings. The fifth node lists various configuration guides but does not include specific instructions for group synchronization

with an OpenID Connect provider.

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

CustomContextualRelevancy: The retrieval context mentions group synchronization but lacks specific details on configuring `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping` properties in `xwiki.properties`.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.7352272727272727

Individual_scores:

AnswerRelevancy: 0.8125

Faithfulness: 0.7272727272727273

ContextualPrecision: 1.0

ContextualRecall: 0.6363636363636364

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.81 because the main content addresses the question about configuring file upload restrictions effectively, but the inclusion of repeated URLs that do not directly answer the question lowers the score.

Faithfulness: The score is 0.73 because the actual output incorrectly states the maximum attachment size as 10 MB, while the retrieval context specifies it as 100GB by default and 32MB for older versions. Additionally, the actual output mentions adding specific lines to restrict attachment size and types, which is not supported by the retrieval context. Lastly, the actual output limits uploadable file types to JPEG, PNG, GIF, and PDF, whereas the retrieval context includes a broader whitelist.

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes, ensuring precise and accurate retrieval.

ContextualRecall: The score is 0.64 because the retrieval context covers the steps for configuring the maximum attachment size in detail (sentences 6-11), but it does not mention configuring MIME types in the `xwiki.properties` file (sentences 2-5).

CustomContextualRelevancy: The retrieval context partially covers the steps to change the maximum attachment size but lacks specific details on configuring MIME types and setting the size to 10 MB.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_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 provides some relevant information, but it includes multiple irrelevant URLs that do not address the specific question about denying script rights to a domain administrator.

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 discusses 'image links and references', which is not relevant to the question about denying script rights to a space administrator.

ContextualRecall: The score is 0.00 because the expected output sentence does not match any part of the nodes in retrieval context, which primarily discuss image references, LDAP authentication, and form utilization in XWiki.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_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 provided URLs are general resources and do not specifically address the configuration of permissions for teams in XWiki. However, the answer does contain some relevant information related to XWiki, which is why the score is not lower.

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 discusses actions in XWiki but not the specific steps needed to configure access control for teams, the second node talks about macros like 'Image' and 'Menu' which are unrelated, and the third node explains XWiki Syntax 2.1 without addressing access control 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 information about creating groups for each team, setting wiki-wide rights, setting area-level rights, or inheriting rights for child pages.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.33999999999999997

Individual_scores:

AnswerRelevancy: 0.7

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.70 because while the response provides some relevant information, it includes multiple statements about managing notifications in Windows 11 that do not directly address the issue with the 'Notification d'alerte' extension.

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 how to get a notification RSS feed and various settings related to notifications, but it does not address enabling notifications in the wiki or setting `notifications.enabled` in `xwiki.properties`. The second node is about overriding default notification templates and installation instructions for extensions, which is not relevant to enabling notifications in the wiki. The third node describes the 'Clear All' link and how notifications are grouped, but it does not provide information on enabling notifications in the wiki. The fourth node explains custom notification filters and the watch feature, which is unrelated to enabling notifications in the wiki. The fifth node provides a description and various settings for notifications, but it does not mention enabling notifications in the wiki or setting `notifications.enabled` in `xwiki.properties`.

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

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.7223809523809523

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.75

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is highly relevant and addresses the issue of not receiving notifications, but the inclusion of an unrelated URL slightly detracts from its overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect faithfulness!

ContextualPrecision: The score is 0.33 because the relevant node (third node) is ranked lower than three irrelevant nodes. The first node discusses getting a notification RSS feed and settings for notifications, the second node is about overriding default notification templates and installation instructions, and the fourth node is about clearing notifications and grouping similar notifications. These should be ranked lower than the third node, which mentions 'Watched Entities' and 'Follow a user', directly relevant to following pages or users to receive notifications.

ContextualRecall: The score is 0.75 because several key aspects of the expected output, such as following pages or users for notifications, are well-supported by nodes in the retrieval context. However, specific details like defining the property `notifications.enabled` in `xwiki.properties` and the 'network' tab in the user profile are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context contains information about enabling notifications, following users, and notification settings, but lacks details on following pages, specific instructions for setting `notifications.enabled`, 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: 0.6

Individual_scores:

AnswerRelevancy: 0.7

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.70 because while the response contains some relevant information, it also includes multiple irrelevant links and headings that do not directly address the question about enabling notifications for one's own 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 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses how to get an RSS feed of notifications and does not address notifications for one's own activity. Similarly, the second node lists various notification settings and filters but does not specifically mention how to enable notifications for one's own actions.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the nodes in the retrieval context, providing a clear and accurate match.

CustomContextualRelevancy: The retrieval context mentions default notification settings and filtering options but lacks details about disabling the filter in 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7138095238095238

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 0.5

CustomContextualRelevancy: 0.7

Reasons:

AnswerRelevancy: The score is 0.79 because the answer provides some relevant information about disabling notifications, but includes multiple irrelevant URLs that do not address the specific question asked.

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

ContextualPrecision: The score is 0.58 because the first node discusses getting a notification RSS feed and the settings menu, which is not relevant to disabling notifications or auto-watched pages.

Additionally, the fourth node provides general information about different ways to receive notifications, and the fifth node discusses overriding default notification templates and installation instructions, both of which are also not relevant. However, the second and third nodes are correctly ranked higher as they directly address modifying auto-watch settings and managing notification filters.

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 remaining sentences about being notified of changes on your work and deleting filters created automatically are not explicitly supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions auto-watched pages and notification settings but lacks details about modifying auto-watch settings and removing custom filters.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.46461538461538465

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is highly relevant and addresses the main question about configuring XWiki for automatic user account activation using OpenID Connect. However, it includes an irrelevant link to XWiki RESTful API documentation, which slightly detracts from its overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context provide general information about OpenID Connect Authenticator, container authentication, or various authentication methods, but none address the specific configuration needed to automatically activate user accounts. For example, the first node provides general information about the OpenID Connect Authenticator, including installation and developer details, but does not address the specific configuration needed to automatically activate user accounts.

ContextualRecall: The score is 0.40 because while the 3rd node in the retrieval context mentions 'OpenId Connect Authentication' and 'xwiki.cfg configuration file', aligning with adjusting 'xwiki.properties' for OpenID Connect, it does not mention adding or updating the specific property 'oidc.enableUser=true', nor does it mention saving the file and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context does not contain information about adjusting the 'xwiki.properties' file or setting the 'oidc.enableUser' property.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.43333333333333334

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the answer addresses the main question about customizing user profile fields from an OpenID Connect provider in XWiki, but it includes irrelevant URLs that do not contribute to solving the specific issue.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.33 because the relevant node (third node) is ranked lower than several irrelevant nodes. For example, the first node discusses 'various authentication methods for the XWiki RESTful API,' and the second node provides information about 'the OpenID Connect Authenticator extension for XWiki,' neither of which address mapping user profile fields from an OpenID Connect provider. However, the third node, which mentions the `xwiki.properties` file and configuring OpenID Connect endpoints, is relevant.

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

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.65109243697479

Individual_scores:

AnswerRelevancy: 0.9411764705882353

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.21428571428571427

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant to configuring group synchronization between XWiki and an OpenID Connect provider, but it includes a minor irrelevant statement about creating new users in XWiki.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.50 because the second node, which is relevant and mentions the default group synchronization and the use of the `oidc.groups.mapping` property, is ranked second. However, irrelevant nodes like the first node, which provides general information about the OpenID Connect Authenticator and its installation, and the third node, which provides general information about the OpenID Connect Authenticator, are ranked higher than relevant nodes.

ContextualRecall: The score is 0.21 because the retrieval context nodes mention configuring properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, but they lack specific steps, examples, and details about locating and editing the `xwiki.properties` file, as well as saving and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions configuring group synchronization in the xwiki.properties file and the oidc.groups.mapping property but lacks details on setting oidc.groups.claim, oidc.groups.separator, and saving the file.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_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 answer addresses the question about restricting file uploads based on type and size, but it includes irrelevant URLs that do not directly relate 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 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'various macros and functionalities in XWiki' which does not address the configuration of MIME types or attachment size limits. Similarly, the second node is about 'customizing the PDF cover in XWiki,' which is unrelated to setting MIME types or attachment size limits.

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

CustomContextualRelevancy: The retrieval context does not contain information about configuring MIME types or setting maximum upload size.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_029_result.json

Overall_score: 0.45999999999999996
Individual_scores:
AnswerRelevancy: 0.0
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 1.0
CustomContextualRelevancy: 0.3
Reasons:

AnswerRelevancy: The score is 0.00 because the provided output contains multiple irrelevant statements about Microsoft Teams and unrelated XWiki documentation links, which do not address the question about refusing script rights to a space administrator in XWiki.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node 'discusses additional rules for rights but does not address the specific question about denying script rights to an administrator.' Additionally, the second node 'provides details on executing tasks with user rights but does not mention the possibility of denying script rights to an administrator.'

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

CustomContextualRelevancy: The retrieval context mentions that an admin right implies other rights and cannot be overridden, but it does not explicitly state that admin rights cannot be refused.

Question_language: fr
Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.3547619047619047

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 0.9166666666666666

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.86 because the answer is mostly relevant and provides useful information, but it includes a URL that does not directly address the specific question about configuring team spaces in XWiki.

Faithfulness: The score is 0.92 because the actual output incorrectly claims to select 'Espaces' in the 'Gestion des droits' menu, while the retrieval context states that the notion of space has been removed from the XWiki UI.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to configuring access controls for team spaces. For example, the first node explains how pages and spaces are created and managed in XWiki, which does not address the input query. Similarly, the second node lists various macros and their functionalities, which is unrelated to the query.

ContextualRecall: The score is 0.00 because none of the steps or details mentioned 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.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.5352380952380953

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.14285714285714285

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.83 because the answer is mostly relevant to the question about authenticating users with access tokens, but it includes multiple irrelevant URLs that do not directly address the query.

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

ContextualPrecision: The score is 0.50 because the relevant node (rank 2) about configuring 'Authorized Applications' is ranked higher than some irrelevant nodes. However, the first node discusses claims at the UserInfo endpoint, which is irrelevant to authenticating users with access tokens and should be ranked lower. Similarly, nodes ranked 3, 4, and 5 provide general security information, cookie protection, and user profile visibility, which are all unrelated to the input query and should also be ranked lower.

ContextualRecall: The score is 0.14 because the retrieval context only partially supports the expected output, specifically regarding the configuration of 'Authorized Applications' (sentence 2), while missing key details about token-based authentication, JWT specifics, and other necessary configurations.
CustomContextualRelevancy: The retrieval context mentions 'Authorized Applications' and JWT tokens but lacks details on enabling the token authenticator, specific claims, and usage instructions.
Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_032_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 answer is perfectly relevant and directly addresses the question about authenticating users with access tokens. Great job!
Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!
ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'group synchronization and customization for OpenID Connect', which is not directly related to authenticating users with access tokens. The second node talks about 'protection levels for cookies and security strategies', which does not address the authentication of 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 specific steps for configuring the LLM application (BETA) extension.
Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_033_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 answer provides relevant information about authenticating users with access tokens, but includes multiple URLs that are not directly related to the specific question.
Faithfulness: The score is 1.00 because there are no contradictions. Great job staying faithful to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'properties of a collection object and methods for managing collections in a wiki', which is unrelated to user authentication using access tokens.

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

CustomContextualRelevancy: The retrieval context does not contain information about authenticating users with access tokens, JWT claims, or configuring the authenticator as described in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_gemma2_9B_Q4

File: qa_001_result.json

Overall_score: 0.8095238095238095

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.71 because while the answer addresses the main issue, it includes several headings that do not provide actionable information, making parts of the response irrelevant.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.33 because the relevant node mentioning the 'notifications.enabled' setting in 'xwiki.properties' is ranked third, while irrelevant nodes discussing topics like 'getting a notification RSS feed and various settings' (first node) and 'overriding default notification templates' (second node) are ranked higher.

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

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.5757142857142857

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.42857142857142855

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully 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. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.25 because the relevant node, which mentions 'notifications.enabled' and 'notifications.emails.enabled' settings in 'xwiki.properties', is ranked fourth. Irrelevant nodes, such as the first node discussing the 'Clear All' link, default notification settings, and grouping notifications, should be ranked lower.

ContextualRecall: The score is 0.43 because while there are mentions of enabling notifications and the alert menu (nodes 3 and 5 in the retrieval context), there are no references to subscribing to pages or users, following a user, receiving notifications about user events, or the 'network' tab in the user profile.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties but lacks details on subscribing to pages or users and watching locations.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.44000000000000006

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and there are no irrelevant statements in the actual output. 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 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses getting a notification RSS feed and general notification settings but does not address enabling notifications for one's own actions. The second node lists options for customizing notification emails and creating global notification filters but does not mention enabling notifications for one's own actions. The third node is about enabling and disabling the email feature and customizing live email notifications grace time, which is not relevant to enabling notifications for one's own actions. The fourth node lists different ways to be notified in XWiki, such as through the Notifications Application, Share Page by mail, RSS Feeds, and Mentions, but does not address enabling notifications for one's own actions. The fifth node discusses overriding default notification templates and installation instructions for extensions, which is not relevant to enabling notifications for one's own actions.

ContextualRecall: The score is 0.00 because the retrieval context does not mention that users don't receive notifications about their own activity unless explicitly targeted, nor does it mention the ability to disable this filter in the 'Advanced filtering options' section of the user's notification settings.

CustomContextualRelevancy: The retrieval context mentions default notification settings and the ability to enable/disable notifications, but does not specify disabling the filter for one's own activity or the 'Advanced filtering options' section.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6266666666666667

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

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

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.33 because the relevant node, which mentions 'Watch and custom filters' and is relevant to modifying notification settings, is ranked third. However, irrelevant nodes discussing enabling/disabling the notification module (first node), getting a notification RSS feed (second node), 'Clear All' link (fourth node), and choosing default frequency (fifth node) are ranked higher than the relevant node.

ContextualRecall: The score is 0.50 because the retrieval context partially supports the expected output, specifically regarding notifications for pages users created or updated (sentence 1), but lacks information on 'Automatic page watching' and deleting notification filters (sentences 3 and 4).

CustomContextualRelevancy: The retrieval context mentions settings for notifications and auto-watched mode but lacks details on major modifications, notification filters, and specific user profile settings.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_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 provides relevant information about configuring OpenID Connect authentication, but it includes irrelevant links related to the XWiki RESTful API and access rights management, which are not directly related to the question.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect faithfulness!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node 'describes the OpenID Connect Authenticator extension but does not provide any information on configuring user accounts to be automatically enabled.' The second node 'discusses container authentication and mentions OpenID Connect briefly but does not provide the specific configuration needed for enabling user accounts.' The third node 'discusses various authentication methods including OpenID Connect but does not mention the specific configuration required to enable user accounts automatically.' The fourth node 'is about setting access rights for terminal pages and does not relate to configuring OpenID Connect or enabling user accounts.' The fifth node 'is about logging in and recovering usernames or passwords, which is not relevant to configuring OpenID Connect to enable user accounts automatically.'

ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.
CustomContextualRelevancy: The retrieval context does not contain information about adjusting the `xwiki.properties` file or setting the `oidc.enableUser` property to `true`.
Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_006_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 is mostly relevant and provides useful information, but it includes links that are not directly related to customizing user profile fields from an OpenID Connect provider, which slightly detracts from its 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 0.00 because all nodes in the retrieval context are irrelevant. The first node provides general information about the OpenID Connect Authenticator, but does not mention anything about customizing user profile fields or mapping additional fields. The second node discusses various authentication methods for the XWiki RESTful API, but does not address customizing user profile fields or mapping additional fields from OpenID Connect. The third node explains the configuration of the OIDC authenticator in XWiki, but does not provide information on mapping additional user profile fields or customizing user profile fields. The fourth node details the features available on the global Administration page, including user management, but does not mention anything about customizing or mapping user profile fields from OpenID Connect. The fifth node provides information on various authentication methods, including OpenID Connect, but does not discuss customizing user profile fields or mapping additional fields.

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

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

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_007_result.json

Overall_score: 0.49247252747252757
Individual_scores:
AnswerRelevancy: 0.9230769230769231
Faithfulness: 0.875
ContextualPrecision: 0.25
ContextualRecall: 0.21428571428571427
CustomContextualRelevancy: 0.2
Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and helpful, but it includes an irrelevant link to the XWiki RESTful API documentation, which is not directly related to configuring group synchronization between XWiki and an OpenID Connect provider.

Faithfulness: The score is 0.88 because the actual output incorrectly states that adding the claim `xwiki_groups` to the `oidc.userinfoclaims` property enables user info claims, while the retrieval context specifies that it enables default group synchronization.

ContextualPrecision: The score is 0.25 because the relevant node providing specific information on configuring group synchronization is ranked fourth, while irrelevant nodes discussing general OpenID Connect details, installation instructions, and customization options are ranked higher (first, second, third, fifth, and sixth).

ContextualRecall: The score is 0.21 because while the retrieval context mentions configuring group synchronization and mapping groups using the xwiki.properties file (supporting sentences 1, 3, and 6), it lacks details on locating and editing the file, specific examples, and instructions on saving and restarting (unsupporting sentences 2, 4, 5, 7, 8, 9, 10, 11, and 12).

CustomContextualRelevancy: The retrieval context mentions group synchronization and the `oidc.groups.mapping` property but lacks detailed steps and specific properties like `oidc.groups.claim` and `oidc.groups.separator` found in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.7735177865612648

Individual_scores:

AnswerRelevancy: 0.9130434782608695

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.454545454545453

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.91 because the answer is mostly relevant and provides useful information about configuring attachment settings, but it includes some irrelevant details about XWiki syntax and access to configuration files that do not directly address the input 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 1.00 because the relevant node, which mentions that 'As an administrator you can set limits on the maximum size of an attachment,' is ranked first, while all irrelevant nodes are ranked lower.

ContextualRecall: The score is 0.45 because while the retrieval context provides detailed steps for setting the maximum attachment size (Step 2), it lacks information on configuring allowed mimetypes and editing the `xwiki.properties` file (Step 1).

CustomContextualRelevancy: The retrieval context contains detailed steps for setting maximum attachment size and mentions mimetype restrictions, but lacks specific instructions for configuring the `xwiki.properties` file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.40095238095238095

Individual_scores:

AnswerRelevancy: 0.9047619047619048

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.1

Reasons:

AnswerRelevancy: The score is 0.90 because the answer is mostly relevant and provides useful information on how to deny script rights to a space administrator. However, it includes links to unrelated documentation 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. For example, the first node discusses methods for checking access rights and does not address the specific question of denying script rights to a space administrator. Similarly, the second node talks about the usage of the `checkAccess()` method without providing information on denying script rights to a space administrator.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about rights implied by admin rights or their denial, focusing instead on access checks and methods like `checkAccess()` and `hasAccess()`.

CustomContextualRelevancy: The retrieval context does not mention that admin rights cannot be denied, which is the main point of the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.4589473684210527

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.89 because the answer mostly addresses the question about configuring access rights in XWiki, but includes URLs that are not directly relevant to the specific task of configuring access rights.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.20 because the relevant node (rank 5) discussing 'how permissions are set for groups and users at different scopes' is ranked lower than irrelevant nodes. For instance, the first node 'explains how pages and spaces are created and managed in XWiki' but does not address configuring access rights for teams, and should be ranked lower.

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 mentions setting rights and creating spaces but lacks specific details on creating groups for each team and inheriting rights for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.3875

Individual_scores:

AnswerRelevancy: 0.9375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.94 because the response is highly relevant and addresses the issue with the notification bell icon, but the statement about not having access to the XWiki instance is unnecessary.

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. For example, the first node 'discusses how to get a notification RSS feed and the settings related to notifications but does not address enabling notifications in the wiki settings.' This pattern continues with subsequent nodes, such as the second node which 'talks about overriding default notification templates and installation instructions but does not mention enabling notifications in the wiki settings.'

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications via the `notifications.enabled` setting in `xwiki.properties` or setting it to `true`.

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

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_012_result.json

Overall_score: 0.275

Individual_scores:

AnswerRelevancy: 0.375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.38 because the provided URL is not directly related to notifications in XWiki, which is the main concern in the input.

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. For example, the first node discusses 'various XWiki applications like Meeting Application, Forum Application, and Ideas Application,' which are not relevant to enabling notifications. Similarly, the second node talks about 'access rights and executing tasks with specific user rights,' which is unrelated to 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.

CustomContextualRelevancy: The retrieval context does not contain any information about enabling notifications or following pages and users.

Question_language: de
Expected_answer_language: de
Answer_language: en

File: qa_013_result.json

Overall_score: 0.3230769230769231
Individual_scores:
AnswerRelevancy: 0.6153846153846154
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.62 because while the output may contain some relevant information, it is cluttered with multiple irrelevant URLs that do not address the specific question about enabling notifications for one's own activity.

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 unrelated to the input. For example, the first node is about 'token generation and JWT claims', which is irrelevant to notifications for user activity. Similarly, the second node discusses 'database migration and invalid extensions', which also does not relate to user notification settings.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention notifications about user activity, related settings, or 'Erweiterten Filteroptionen'.

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

Question_language: de
Expected_answer_language: de
Answer_language: en

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 while the response addresses the question about disabling notifications, it includes multiple irrelevant URLs that do not provide the requested information, reducing the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions. 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 'Verbatim' and 'Quotations' in XWiki Syntax, which is unrelated to disabling notifications. Similarly, the second node provides information on 'JWT Token Authenticator' for the LLM AI Application, which does not pertain to notification settings.

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

CustomContextualRelevancy: The retrieval context does not contain any information about automatic following of pages, notification settings, or custom filters.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.5666666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the answer is mostly relevant and provides useful information, but it includes a statement about the inability to provide a direct answer, which does not directly address the input query.

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 provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed. The second node describes the purpose of the authenticator but lacks details on automatic user activation. The third node lists various configuration guides but omits the specific property `oidc.enableUser`. The fourth node mentions the `xwiki.properties` file but does not specify the necessary property for automatic user activation.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the 2nd node in the retrieval context. Great job!

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.64

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

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

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect faithfulness!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context provide general information about prerequisites, installation instructions, dependencies, general capabilities, container authentication, endpoints, customization options, and configuration guides, but none address

customizing user profile fields or mapping additional fields. For example, the first node discusses 'prerequisites, installation instructions, and dependencies,' which is not relevant to the input query.
ContextualRecall: The score is 1.00 because all sentences in the expected output can be attributed to the nodes in the retrieval context, indicating a perfect match.

CustomContextualRelevancy: The retrieval context mentions the xwiki.properties file but lacks specific details on customizing user profile fields and the oidc.user.mapping property.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.64

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and directly addresses the question about setting up group synchronization between XWiki and the OpenID Connect provider.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node provides general information about the OpenID Connect Authenticator but does not address the specific process of group synchronization configuration. The second node describes the OpenID Connect Authenticator and its compatibility but does not provide information on configuring group synchronization. The third node mentions container authentication and other methods but lacks details on setting up group synchronization. The fourth node references group synchronization but does not provide the detailed steps or configuration settings required. The fifth node lists various configuration guides and files but does not provide the specific steps or properties needed to configure group synchronization.

ContextualRecall: The score is 1.00 because every sentence in the expected output aligns perfectly with the nodes in the retrieval context, providing a comprehensive and accurate guide for setting up group synchronization between XWiki and an OpenID Connect provider.

CustomContextualRelevancy: The retrieval context mentions group synchronization but lacks detailed steps and specific 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: 0.7868831168831169

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 0.75

ContextualPrecision: 1.0

ContextualRecall: 0.7272727272727273

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.86 because the majority of the answer is relevant to configuring attachment settings in XWiki, but it includes links to an extension and a user guide that are not directly related to the specific query.

Faithfulness: The score is 0.75 because the actual output incorrectly mentions configuring 'attachment.download.whitelist' and setting the maximum size of an attachment to 10 MB, which are not supported by the retrieval context.

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes, ensuring precise and accurate retrieval.

ContextualRecall: The score is 0.73 because while the retrieval context nodes provided detailed steps for configuring MIME types and maximum upload size (supporting sentences 2, 3, and 4), the introductory statement and headings (sentences 1, 5, and 6) were not directly supported by any specific node in the retrieval context.

CustomContextualRelevancy: The retrieval context includes steps to configure maximum attachment size and mentions MIME type restrictions, but it lacks specific details on configuring MIME types in the xwiki.properties file and setting the maximum size to 10 MB.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_result.json

Overall_score: 0.23333333333333334

Individual_scores:

AnswerRelevancy: 0.16666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.17 because the provided URLs and statements are irrelevant to the question about denying script rights to a space administrator. The response did not address the specific query.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input question. For instance, the first node contains 'image links and references to various projects', the second node includes 'technical details about Kerberos authentication and LDAP configuration', and the third node discusses 'mail templates for resetting passwords and finding usernames'. None of these contexts address the question about denying script rights to a domain administrator in XWiki.

ContextualRecall: The score is 0.00 because the expected output sentence does not match any part of the provided retrieval context, which discusses images, LDAP authentication, and form utilization on XWiki, but does not mention admin rights or their implications.

CustomContextualRelevancy: The retrieval context does not contain any information about the admin rights or their implications in XWiki.

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 answer provides some relevant information about configuring permissions in XWiki, but includes several irrelevant links that do not directly address the specific configuration question.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses actions like creating children pages and changing page access rights, but it does not address the specific requirements of configuring access for teams. The second node provides information about various macros and their functionalities in XWiki, which is unrelated to configuring team-specific access controls. Similarly, the third node explains aspects of XWiki Syntax 2.1, which does not provide information on setting up team-specific access controls.

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

CustomContextualRelevancy: The retrieval context does not contain specific steps for creating groups, setting wiki-wide rights, area-level rights, or inheriting rights for subpages as outlined in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_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 there are no irrelevant statements.

Faithfulness: The score is 1.00 because there are no contradictions. Great job staying faithful to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses how to get a notification RSS feed and the settings for notifications, but it does not address how to enable notifications in the wiki.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about ensuring notifications are activated in the wiki or provide any information about defining the property `notifications.enabled` in `xwiki.properties`.

CustomContextualRelevancy: The retrieval context does not contain any information about setting the property `notifications.enabled` in `xwiki.properties` to `true`.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.5044444444444445

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.89 because the response is mostly relevant to the issue of not receiving notifications, but the inclusion of a URL that does not directly address this issue slightly lowers the score.

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

ContextualPrecision: The score is 0.33 because the relevant node, which mentions 'Watched Entities' and 'Follow a user', is ranked third. Irrelevant nodes, such as the first node discussing 'getting a notification RSS feed and settings' and the second node about 'overriding default notification templates and installation instructions', should be ranked lower.

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 mentions enabling notifications and following users, but lacks detailed steps on setting `notifications.enabled` to `true` and specifics on how to follow pages or users.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.7066666666666668

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.83 because the answer provides useful information about enabling notifications, but includes irrelevant URLs that do not directly address the query about notifications for one's own actions.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.50 because the relevant node ranked 2nd mentions 'Filter your own events', which is pertinent to receiving notifications for one's own activity. However, irrelevant nodes ranked 1st, 3rd, 4th, and 5th discuss topics such as 'notification RSS feed', 'various notification preferences', 'overriding default notification templates', and 'various ways to be notified in XWiki', which do not address the specific issue.

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

CustomContextualRelevancy: The retrieval context mentions filtering own events and notification settings but lacks details on default behavior and advanced filtering options.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.61

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 0.6666666666666666

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.50 because the provided URLs are not directly relevant to the input, which specifically asks how to disable unwanted notifications. However, the response does partially address the issue, hence the score is not lower.

Faithfulness: The score is 0.67 because the actual output incorrectly claims that users can disable notifications for pages they haven't created or updated, whereas the retrieval context specifies that users only receive notifications when they are mentioned or when actions are performed on pages they created or updated.

ContextualPrecision: The score is 0.33 because the relevant node, which discusses 'Auto Watch' and 'Default Watch Settings,' is ranked third, while the first and second nodes are irrelevant and should be ranked lower as they do not address how to disable notifications or the automatic activation of notifications.

ContextualRecall: The score is 0.75 because the retrieval context nodes support the general idea of notifications and settings, but they do not specifically mention being notified of changes on your work.

CustomContextualRelevancy: The retrieval context covers auto-watch settings, notification preferences, and custom filters but lacks specific details on modifying existing filters and the exact notification of changes on user work.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.46749999999999997

Individual_scores:

AnswerRelevancy: 0.9375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and addresses the main question about configuring OpenID Connect for automatic user activation in XWiki. However, it includes a minor irrelevant statement about the XWiki RESTful API, which slightly detracts from the overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context provide general information about the OpenID Connect Authenticator or related topics, but none address the specific configuration needed to automatically activate user accounts. For example, the first node provides general information about the OpenID Connect Authenticator, including its developer, type, and license, but does not address the specific configuration needed to automatically activate user accounts.

ContextualRecall: The score is 0.40 because while the 3rd node in the retrieval context mentions 'automatic synchronization' and 'OpenID Connect Authenticator Extension,' which aligns with activating user accounts on first login, there is no specific mention of adding or updating properties, the property `oidc.enableUser`, or saving the file and restarting the instance.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the `xwiki.properties` file or setting the `oidc.enableUser` property.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.4733333333333344

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the answer is mostly relevant but includes repeated irrelevant URLs that do not address the specific question about customizing user profile fields from the OpenID Connect provider in XWiki.

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 (third node) is ranked lower than multiple irrelevant nodes. For example, the first node discusses 'various authentication methods for the XWiki RESTful API, including OIDC, but does not address customizing user profile fields or mapping additional fields', and the second node provides 'general information about the OpenID Connect Authenticator for XWiki, including installation and compatibility details, but does not mention how to map additional user profile fields'.

ContextualRecall: The score is 0.20 because only the mention of 'xwiki.properties' and 'OIDC' in the 3rd node in the retrieval context supports the expected output, while key details such as 'oidc.user.mapping' property, examples of mapping properties, replacing 'myxproperty1' and 'myxproperty2', and saving and restarting XWiki are missing.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6989966555183946

Individual_scores:

AnswerRelevancy: 0.9565217391304348

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.5384615384615384

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.96 because the answer is highly relevant and addresses the main question about configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes a minor irrelevant statement about creating new users, which slightly detracts from the overall relevance.

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

ContextualPrecision: The score is 0.50 because the relevant node that mentions the configuration of group synchronization in the `xwiki.properties` file is ranked second, while irrelevant nodes that provide general information, prerequisites, and unrelated configuration details are ranked higher, such as the first node which mainly provides links to prerequisites and installation instructions.

ContextualRecall: The score is 0.54 because while the retrieval context covers key properties like 'oidc.groups.claim', 'oidc.groups.separator', and 'oidc.groups.mapping' along with examples, it lacks explicit instructions on locating, opening, modifying, and saving the 'xwiki.properties' file, as well as restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions configuring group synchronization in xwiki.properties and mapping groups, but lacks detailed steps and examples found in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_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 answer is mostly relevant but includes a URL that does not directly address the specific question about limiting file uploads to images or PDFs of a maximum size of 10 MB.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to configuring file upload restrictions. For example, the first node discusses 'various macros and their functionalities,' the second node explains 'how to customize the PDF cover,' and the third node describes 'how to create a new page,' none of which 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 about configuring MIME types or setting attachment size limits.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.9349999999999999

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.88 because the answer is mostly relevant and provides useful information, but it includes URLs that do not directly address the specific question of how to refuse script rights to a space administrator.

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, ensuring precise and accurate retrieval.

ContextualRecall: The score is 1.00 because the expected output perfectly matches the provided context. Great job!

CustomContextualRelevancy: The retrieval context explains that rights implied by admin rights cannot be overridden, aligning with the expected answer that admin rights cannot be refused. However, it does not explicitly state that XWiki does not support this feature.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_030_result.json

Overall_score: 0.22260869565217392

Individual_scores:

AnswerRelevancy: 0.9130434782608695

Faithfulness: 0.2

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.91 because the answer is mostly relevant and provides useful information, but it includes a URL that does not directly address how to configure permissions in XWiki, which slightly detracts from its overall relevance.

Faithfulness: The score is 0.20 because the actual output repeatedly mentions controlling permissions and creating spaces, while the retrieval context clearly states that the notion of space has been removed from the UI.

ContextualPrecision: The score is 0.00 because all the nodes in the retrieval context are irrelevant to configuring access rights for team spaces. For example, the first node explains the concept of spaces and sub-pages in XWiki but does not provide any information on how to set access rights for teams. The second node describes various macros and their uses in XWiki, which is unrelated to configuring access rights for team spaces. Additionally, the third node discusses the parent/child relationship and nested spaces in XWiki, but it does not provide any steps or guidelines on setting access rights for different teams.

ContextualRecall: The score is 0.00 because none of the steps or details 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.

Question_language: fr
Expected_answer_language: fr
Answer_language: en

File: qa_031_result.json

Overall_score: 0.5215126050420168
Individual_scores:
AnswerRelevancy: 0.7647058823529411
Faithfulness: 1.0
ContextualPrecision: 0.5
ContextualRecall: 0.14285714285714285
CustomContextualRelevancy: 0.2
Reasons:

AnswerRelevancy: The score is 0.76 because while the answer partially addresses the question about authenticating users with access tokens, it includes multiple irrelevant URLs that do not directly address the query.

Faithfulness: The score is 1.00 because there are no contradictions present. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.50 because the first node discusses claims at the UserInfo endpoint, which is not about authenticating users with access tokens, and should be ranked lower. The third node is about permissions and JAR types, which is also irrelevant and should be ranked lower. The fourth node discusses 'remember me' cookie functionality and Kerberos SSO Authentication, which are not related to access token authentication and should be ranked lower. The fifth node is about user profile visibility and custom claims, which does not address user authentication with access tokens and should be ranked lower. However, the second node correctly explains configuring authorized applications, which is relevant to the process.

ContextualRecall: The score is 0.14 because only sentence 2 in the expected output can be attributed to the 2nd node in the retrieval context, while the rest of the sentences have no corresponding nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions the 'Authorized Applications' configuration but lacks details on JWT claims, token generation, and specific authenticator setup steps.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_032_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 answer is fully relevant and directly addresses the question 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 0.00 because all nodes in the retrieval context are irrelevant to access token-based authentication. For instance, the first node discusses 'group synchronization and

customization for OpenID Connect', which is not related to the input query. Similarly, the second node mentions 'protection levels for cookies and security strategies for failed logins', which also do not pertain to the input query.

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

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication for the LLM application (BETA) extension, JWT tokens, or the specific configuration steps mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_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 provided URLs do not directly address the question about authenticating users with access tokens, making the response partially irrelevant.

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 'properties of a collection resource and methods for managing collections', which is unrelated to user authentication with access tokens. Similarly, the second node discusses 'access tokens and claims at the UserInfo endpoint' but does not provide 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 about authenticating users with access tokens, JWT, or the specific steps and requirements mentioned in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_llama3_1_402b

File: qa_001_result.json

Overall_score: 0.7923809523809524

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9285714285714286

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.7

Reasons:

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

Faithfulness: The score is 0.93 because the actual output incorrectly states that the Alerts Application must be installed and enabled for the Notifications Application to display the Bell icon, whereas the retrieval context only mentions the Notifications Application contributing content to the Alert menu.

ContextualPrecision: The score is 0.33 because the relevant node, which mentions the `notifications.enabled` setting in `xwiki.properties`, is ranked third. However, irrelevant nodes discussing 'getting a notification RSS feed and settings for notifications' (first node), 'overriding default notification templates and installation instructions' (second node), 'the 'Clear All' link and grouping of notifications' (fourth node), and 'the notifications menu and the bell icon' (fifth node) are ranked higher or close to the relevant node.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports the expected output. Great job!

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties but does not explicitly state setting 'notifications.enabled' to 'true'.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.5623809523809524

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9333333333333333

ContextualPrecision: 0.25

ContextualRecall: 0.42857142857142855

CustomContextualRelevancy: 0.2

Reasons:

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

Faithfulness: The score is 0.93 because the actual output incorrectly claims that you can check if the notification module is enabled by checking the `notifications.enabled` property in the `xwiki.properties` file, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.25 because the relevant node (fourth node) is ranked lower than several irrelevant nodes. For example, the first node discusses the 'Clear All' link, default notification settings, and grouping of notifications, but does not address enabling notifications or subscribing to pages/users. The second node lists dependencies for the notifications extension but does not provide information on enabling notifications or subscribing to pages/users. The third node describes various ways to receive notifications but does not explain how to enable notifications or subscribe to pages/users. The fifth node describes the 'What's New' application but does not provide information on enabling notifications or subscribing to pages/users.

ContextualRecall: The score is 0.43 because while the retrieval context supports enabling notifications and watching pages or wikis, it lacks information on subscribing to pages or users, following users, and the 'network' tab in user profiles.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties and the alert menu, but lacks details on subscribing to pages or users and viewing the network tab in the user profile.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.5349999999999999

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.875

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and there are no irrelevant statements.

Faithfulness: The score is 0.88 because the actual output claims enabling notifications for your own actions in XWiki will enable notifications for all users for the selected notification types, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context discuss various aspects of notifications in XWiki but none address enabling notifications for one's own actions. For example, the first node discusses how to get a notification RSS feed and general notification settings, while the second node lists various notification settings and customization options. These irrelevant nodes should be ranked lower.

ContextualRecall: The score is 0.50 because the retrieval context only partially supports the expected output. Specifically, the 1st sentence in the expected output is supported by the 1st node in the retrieval context, but the 2nd sentence about disabling the filter in 'Advanced filtering options' is not mentioned in any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions default notification settings and the ability to enable/disable notifications, but it lacks specific details about disabling the filter for one's own activity and the 'Advanced filtering options' section.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5951515151515152

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9090909090909091

ContextualPrecision: 0.3666666666666667

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Reasons:

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

Faithfulness: The score is 0.91 because the actual output incorrectly states that you need to use the 'Own event filter' switch to filter out events triggered by yourself, while the retrieval context indicates that such notifications are not received by default.

ContextualPrecision: The score is 0.37 because irrelevant nodes like the first node ('enabling and disabling the notification module and email notifications') and the second node ('getting a notification RSS feed and customizing notification settings') are ranked higher than relevant nodes like the third node ('Watch and custom filters') and the fifth node ('Choosing the default auto-watched mode').

ContextualRecall: The score is 0.50 because only the first two sentences in the expected output are supported by the 2nd node in the retrieval context, while the last two sentences are not explicitly mentioned in any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions settings and filters but lacks specific details about automatic page watching and deleting filters.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_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 configuration of OpenID Connect authentication and enabling user accounts. However, it includes irrelevant details about the XWiki RESTful API and Access Rights, which do not directly pertain to the specific configuration issue.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node 'does not provide any information about configuring user accounts to be automatically enabled upon first login,' and the second node 'does not provide the specific configuration needed to automatically enable user accounts.'

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 adjusting the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_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 answer is mostly relevant and provides useful information about customizing user profile fields with OpenID Connect in XWiki. However, it includes some irrelevant links that do not directly address the specific 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. For example, the first node 'describes the OpenID Connect authenticator and its general information, but it does not provide any details on customizing user profile fields or mapping additional fields.' The second

node 'discusses various authentication methods available for the XWiki RESTful API, including OIDC, but does not address customizing user profile fields or mapping additional 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 does not contain any information about customizing user profile fields or mapping them in the xwiki.properties file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.6361538461538462

Individual_scores:

AnswerRelevancy: 0.9615384615384616

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.7692307692307693

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.96 because the answer is highly relevant and provides useful information, but it includes an irrelevant link to the XWiki RESTful API documentation, which is not directly related to configuring group synchronization with 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 (rank 4) that provides information on group synchronization is ranked lower than several irrelevant nodes. For example, the first node describes the OpenID Connect Authenticator extension but does not provide specific instructions for configuring group synchronization. Similarly, the second node provides links to various configuration guides but lacks details on group synchronization between XWiki and an OpenID Connect provider.

ContextualRecall: The score is 0.77 because while the retrieval context nodes mention key properties and configurations related to group synchronization, they do not provide specific instructions for locating, opening, saving the `xwiki.properties` file, or restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions group synchronization and properties like `oidc.groups.mapping` but lacks detailed steps on configuring `xwiki.properties` and specific examples.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.7797979797979798

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.45454545454545453

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and provides useful information on configuring attachment settings, but it includes an unnecessary link to XWiki syntax documentation.

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. For example, the first node provides detailed steps on how to set the maximum size of an attachment, which is directly relevant to the expected output. In contrast, the second node discusses moving and renaming attachments, which is not relevant to setting upload restrictions or size limits.

ContextualRecall: The score is 0.45 because while the retrieval context covers the steps for configuring the maximum attachment size (Steps 2.1 to 2.5), it does not mention the steps for ensuring only images or PDF files of a maximum 10MB can be uploaded (Step 1).

CustomContextualRelevancy: The retrieval context partially covers the expected answer. It includes steps to set the maximum attachment size and mentions mimetype restrictions, but lacks specific details on configuring allowed mimetypes and setting the size to exactly 10MB.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_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 is mostly relevant and provides useful information, but it includes URLs that do not directly address the question about denying script rights to a space administrator.

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. For example, the first node discusses methods for checking access rights in code, but does not address the specific question about denying script rights to a space administrator. Similarly, the second node talks about the usage of the checkAccess() method without providing information on denying script rights to a space administrator.

ContextualRecall: The score is 0.00 because the sentence 'This is unfortunately not supported in XWiki, rights that are implied by admin right cannot be denied.' is not found in any part of the provided nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention that rights implied by admin right cannot be denied, which is the core information in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.40076555023923444

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 0.9090909090909091

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.89 because the provided answer is mostly relevant and helpful, but it includes a URL that does not directly address the configuration of access rights in XWiki, which slightly detracts from its overall usefulness.

Faithfulness: The score is 0.91 because the actual output incorrectly states that denying the Edit right for the XWikiAllGroup at the wiki level will not prevent teams from editing other teams' spaces if the Edit right is allowed at the Page & Children level.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node 'discusses the creation of pages and spaces in XWiki but does not provide information on configuring access control for teams,' which is not useful for the input query.

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

CustomContextualRelevancy: The retrieval context mentions setting rights and managing spaces but lacks specific steps for creating groups, setting wiki-wide rights, and inheriting rights for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_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 addresses the issue without any irrelevant information. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'getting a notification RSS feed and settings for notifications,' which does not address enabling notifications in the wiki settings. The second node is about 'overriding default notification templates and installation instructions,' which is not relevant. The third node provides 'a description and various settings related to notifications,' but does not mention enabling notifications. The fourth node discusses 'default frequency of notification emails, customizing templates, and creating filters,' which is also not relevant. Lastly, the fifth node talks about 'clearing notifications and grouping them,' but does not mention 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 the nodes in the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_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 partially addresses the issue of not receiving notifications, but it repeatedly includes an irrelevant URL that does not provide specific information about notification settings or issues in XWiki.

Faithfulness: The score is 1.00 because there are no contradictions 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 discusses 'various applications in XWiki' but does not address notifications. The second node is about 'access rights and executing tasks with specific user rights' without mentioning notifications. The third node focuses on 'renaming of links inside scripting macros' and lacks information on notifications. The fourth node describes 'embedding model and chunking method for documents,' which is unrelated to notifications. The fifth node talks about 'configuring permissions for guest users,' not notifications.

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

CustomContextualRelevancy: The retrieval context does not contain any information related to enabling notifications, 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: 0.3444444444444444

Individual_scores:

AnswerRelevancy: 0.7222222222222222

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.72 because while the answer provides some relevant information, it includes multiple URLs that do not directly address the question about enabling notifications for one's own 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'token generation and JWT claims', which is unrelated to notifications for user activity. Similarly, the second node talks about 'database migration and invalid extensions', which does not pertain to user notification settings.

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

CustomContextualRelevancy: The retrieval context does not contain any information about user notification settings or the ability to disable filters in the user profile.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_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 answer contains multiple irrelevant URLs that do not address the question about disabling notifications. However, the relevant parts of the answer do provide useful information on how to manage notifications, which is why the score is not lower.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'XWiki Syntax and formatting features', the second node is about 'an extension for authenticating users with JSON Web Tokens', the third node lists 'various resources and APIs related to the LLM Application', the fourth node is about 'XWiki development resources and guides', and the fifth node discusses 'database migration and invalid extensions during an upgrade'. None of these 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 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: 0.6346153846153847

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the response is mostly accurate and relevant, but the incorrect statement 'xwiki.authentication.createuser=empty' slightly detracts from the overall quality.

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 retrieval context (rank 4) that mentions the 'xwiki.properties' file and the 'oidc.enableUser' property is ranked lower than irrelevant nodes. The first node provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed. The second node discusses general capabilities and links to configuration guides but does not mention the specific property or file. The third node provides details on container authentication, Active Directory, and LDAP, which are not relevant to the specific configuration required.

ContextualRecall: The score is 1.00 because all sentences in the expected output are well-supported by the nodes in the retrieval context, specifically the 2nd node which mentions 'xwiki.properties' and configuring it for OpenID Connect. Great job!

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.60000000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.6

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and directly addresses the question 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 'Configuration files' and 'xwiki.properties' for customizing user profile fields, is ranked fifth. The first four nodes are irrelevant as they discuss prerequisites, installation instructions, general descriptions, compatibility, container authentication, Active Directory, LDAP, and various endpoints, none of which address the customization of user profile fields or mapping additional properties.

ContextualRecall: The score is 0.60 because while the retrieval context mentions 'customization' and 'xwiki.properties' (supporting sentences 1, 2, and 5 in the expected output), it lacks specific examples and instructions for mapping custom properties (missing details for sentences 3 and 4).

CustomContextualRelevancy: The retrieval context mentions 'xwiki.properties' and customization but lacks specific details on 'oidc.user.mapping' 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: 0.69000000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully 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. Great job maintaining perfect faithfulness!

ContextualPrecision: The score is 0.25 because the relevant node mentioning 'Group synchronization' and providing a link to the relevant section is ranked fourth. Irrelevant nodes, such as the first node which provides general information about the OpenID Connect Authenticator without detailing group synchronization, should be ranked lower.

ContextualRecall: The score is 1.00 because every sentence in the expected output is fully supported by the relevant nodes in the retrieval context, ensuring a comprehensive and accurate guide.

CustomContextualRelevancy: The retrieval context mentions group synchronization and the xwiki.properties file but lacks detailed steps and configuration examples found in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.7725

Individual_scores:

AnswerRelevancy: 0.85

Faithfulness: 0.8125

ContextualPrecision: 1.0

ContextualRecall: 0.7

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.85 because the answer provides some relevant information but includes URLs that do not directly address the specific task of setting attachment restrictions in the wiki.

Faithfulness: The score is 0.81 because the actual output incorrectly states that the maximum size of an attachment is 10 MB, whereas the retrieval context indicates it is 100GB by default and 32MB for XWiki < 10.9RC1. Additionally, the actual output misnames the configuration section and restricts attachment types to images or PDFs, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node provides steps to set limits on the maximum size of an attachment, while the subsequent nodes discuss unrelated topics such as MIME type whitelists, managing documents, creating new pages, and properties of document collections.

ContextualRecall: The score is 0.70 because the retrieval context provides detailed steps for setting the maximum upload size in the *XWikiPreferences* document (related to sentences 6-12 in the expected output) but does not mention configuring MIME types in `xwiki.properties` (related to sentences 2-5 in the expected output).

CustomContextualRelevancy: The retrieval context partially covers the steps to set the maximum attachment size but lacks details on configuring MIME types.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_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 answer provides relevant information about denying script rights to a space administrator, but it includes multiple irrelevant links that do not address the question.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining faithfulness to 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 contains 'a series of image links and references', which does not provide any information relevant to the question. Similarly, the second node discusses 'authentication and LDAP configuration', which is not related to the question. The third node is about 'mail templates for password reset and username recovery', which is also irrelevant. The fourth node provides 'installation instructions for an extension', and the fifth node discusses 'form utilization and adding references', both of which are not related to the question.

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

CustomContextualRelevancy: The retrieval context does not contain any information related to the admin rights or their denial in XWiki, as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_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 answer provides useful information but includes multiple irrelevant links that do not directly address the specific configuration 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 configuring team-specific access and editing rights in XWiki. For instance, the first node discusses actions like creating children pages and changing page access rights, but does not provide specific steps for the required configuration. Similarly, the second node talks about wikis and panels, but does not address team-specific access needs.

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 wiki-wide or area-specific rights, or inheriting rights for child pages as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_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 issue without any irrelevant statements.

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. For example, the first node discusses getting a notification RSS feed and settings for notifications, but does not address enabling notifications in the wiki or setting `notifications.enabled` in `xwiki.properties`. The second node talks about overriding default notification templates and installation instructions for extensions, but does not mention enabling notifications in the wiki or setting `notifications.enabled` in `xwiki.properties`.

ContextualRecall: The score is 0.00 because the retrieval context does not mention activating notifications in the wiki settings or modifying `xwiki.properties`, nor does it provide instructions for setting the `notifications.enabled` property in `xwiki.properties`.

CustomContextualRelevancy: The retrieval context does not mention 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: 0.4833333333333333

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 0.75

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.0

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is highly relevant, but the provided URL does not directly address the issue of not receiving notifications.

Faithfulness: The score is 0.75 because the actual output mentions a 'Filtres' link and a 'Toutes les notifications' link in the notification menu, which are not mentioned in the retrieval context.

ContextualPrecision: The score is 0.33 because the relevant node (third node) is ranked lower than several irrelevant nodes. For example, the first node discusses 'getting a notification RSS feed and settings for notifications,' which is not about enabling notifications or following users/pages. Similarly, the second node is about 'overriding default notification templates and installing extensions,' which is unrelated to the input query.

ContextualRecall: The score is 0.00 because the retrieval context does not provide any information related to ensuring notifications are activated, following pages or users, or the specific processes mentioned in the expected output.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on `xwiki.properties`, the alert menu, and the network tab in the user profile.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.6771428571428572

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 0.8

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.79 because the answer provides some relevant information about notifications, but includes multiple irrelevant URLs that do not directly address the specific question.

Faithfulness: The score is 0.80 because the actual output incorrectly states that users do not receive notifications for their own actions by default, while the retrieval context specifies that users get notifications when mentioned or when actions are performed on pages they created or updated.

Additionally, the actual output incorrectly directs users to click on the 'Settings' link at the bottom of the list of notifications, whereas the retrieval context indicates that the settings link is in the notifications menu.

ContextualPrecision: The score is 0.50 because the relevant node mentioning 'Filter your own events' is ranked second, while the first node discusses general notification settings and the third node lists notification preferences macros, both of which do not address enabling notifications for one's own actions.

ContextualRecall: The score is 1.00 because the expected output perfectly aligns with the nodes in the retrieval context, providing accurate and relevant information.

CustomContextualRelevancy: The retrieval context mentions default notification settings and filtering options, but lacks specific details about disabling the filter in 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.6983333333333335

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 0.8333333333333334

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.75

CustomContextualRelevancy: 0.7

Reasons:

AnswerRelevancy: The score is 0.88 because the main content addresses the query about disabling notifications, but the inclusion of irrelevant links about the Notifications Application, XWiki Syntax, and Notifications features slightly detracts from the overall relevance.

Faithfulness: The score is 0.83 because the actual output incorrectly describes a method to disable notifications by clicking on an icon '...' and selecting 'Désactiver les notifications' from a dropdown menu, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.33 because the relevant node (rank 3) is ranked lower than multiple irrelevant nodes. For example, the first node discusses getting a notification RSS feed and mentions the Notifications Macro, but does not address how to disable automatic notifications or modify notification settings. Similarly, the second node lists various notification-related preferences and

macros but does not provide specific instructions on how to disable automatic notifications or modify notification settings.

ContextualRecall: The score is 0.75 because the retrieval context nodes partially support the expected output, specifically regarding notification settings and custom filters, but do not explicitly mention being notified about changes on your work.

CustomContextualRelevancy: The retrieval context mentions auto watch preferences and custom filters preferences, but lacks details on modifying the auto watch setting and removing filters.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_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 answer is highly relevant and addresses the main question, but it includes an unnecessary mention of the XWiki RESTful API.

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. For example, the first node 'does not provide information on configuring user activation', the second node 'does not address the specific configuration needed for OpenID Connect user activation', and the third node 'does not mention the specific property `oidc.enableUser` or the xwiki.properties file'.

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 specific information about adjusting the `xwiki.properties` file or setting `oidc.enableUser` to `true`.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.5700000000000001

Individual_scores:

AnswerRelevancy: 0.85

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.85 because the response addresses the main question about customizing user profile fields from an OpenID Connect provider in XWiki, but includes several irrelevant URLs 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. For example, the first node discusses 'various authentication methods for XWiki, including OIDC,' but does not provide information about customizing user profile fields or mapping them. Similarly, the second node provides 'general information about the OpenID Connect Authenticator for XWiki,' but does not address customizing user profile fields or mapping them.

ContextualRecall: The score is 1.00 because the retrieval context perfectly aligns with the expected output, covering all key points and steps in detail.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6523076923076923

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.5384615384615384

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and provides useful information on configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes a link to documentation on creating new users in XWiki, which is not directly relevant to the specific question asked.

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

ContextualPrecision: The score is 0.50 because the relevant node ranked 2 mentions 'default group synchronization' and 'relevant properties in the `xwiki.properties` file', which are directly related to the input. However, irrelevant nodes ranked 1, 3, 4, and 5 provide general information, such as 'prerequisites, installation instructions, and dependencies', 'developer, license, and compatibility', 'various endpoints, customization options, and troubleshooting tips', and 'a link to more information about managing users and groups in XWiki', which do not address the specific steps for configuring group synchronization.

ContextualRecall: The score is 0.54 because while the retrieval context covers configuring 'xwiki.properties' and various 'oidc' properties (nodes 2-6), it lacks explicit mentions of locating and opening the file (sentence 1-2), replacing placeholder values (sentences 3, 5, 7), and saving/restarting the instance (sentence 8).

CustomContextualRelevancy: The retrieval context mentions configuring group synchronization in the xwiki.properties file and using properties like oidc.groups.mapping, but it lacks detailed steps and examples present in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3571428571428571

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.79 because the output contains useful information relevant to configuring file upload settings, but it also includes multiple irrelevant URLs that do not address the specific question about restricting file types and sizes.

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. For example, the first node discusses 'various macros and their functionalities' which is unrelated to configuring attachment upload restrictions. The second node explains 'how to customize the PDF cover by overriding the template pdfcover.vm', which does not pertain to setting MIME types or file size limits for attachments. Additionally, the third node describes 'how to create a new page by adding a wanted link', which is not related to configuring attachment 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.

CustomContextualRelevancy: The retrieval context does not contain any information about configuring MIME types or setting the maximum upload size for attachments.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.7466666666666667

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.83 because the answer provides relevant information but includes several irrelevant 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. Great job maintaining accuracy and alignment with the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'default right settler additional policies,' which is not related to denying script rights to an administrator. Similarly, the second node explains 'how to change the current author and execute tasks with specific user rights,' but does not address the specific question. These irrelevant nodes should be ranked lower.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the information in the retrieval context, specifically the 1st node.

CustomContextualRelevancy: The retrieval context explains that rights implied by an admin right cannot be overridden by a denial at a lower level, closely matching the expected answer about admin rights not being able to be refused.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_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 answer is mostly relevant and provides useful information, but it includes a URL that does not directly address the question about configuring access rights in XWiki.

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 configuring access control for team spaces. For instance, the first node discusses the automatic creation of spaces and WebHome pages, the second node lists various macros and their uses, and the third node explains the parent/child relationship and nested pages. None of these nodes provide information on access control or permissions.

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.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.5280926916221034

Individual_scores:

AnswerRelevancy: 0.7647058823529411

Faithfulness: 0.9090909090909091

ContextualPrecision: 0.5

ContextualRecall: 0.16666666666666666

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.76 because the answer provides some relevant information about authenticating users with access tokens, but it includes multiple irrelevant URLs that do not directly address the question.

Faithfulness: The score is 0.91 because the actual output incorrectly claims that an access token is needed to authenticate the user in XWiki, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.50 because the relevant node describing the configuration of 'Authorized Applications' (rank 2) is ranked lower than an irrelevant node discussing various claims accessed via an access token (rank 1). Additionally, other irrelevant nodes discussing general information about a security module (rank 3), cookie protection (rank 4), and user profile rights (rank 5) are also ranked higher than the relevant node.

ContextualRecall: The score is 0.17 because the retrieval context only partially matches the expected output. Specifically, it mentions configuring properties for 'Authorized Applications' (related to sentence 2), but it does not cover key aspects such as 'Token-based authentication for the LLM Application

extension', 'signed JSON Web Tokens (JWT)', or the detailed steps for enabling and using the token authenticator (related to sentences 1, 3, 4, and 5).

CustomContextualRelevancy: The retrieval context mentions 'Authorized Applications' and configuring properties like title, URL, and public key but lacks details on enabling the token authenticator, JWT claims, and token usage.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_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 answer is perfectly relevant and addresses the question without any irrelevant statements. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect faithfulness!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'group synchronization and customization for OpenID Connect,' which is not relevant to authenticating users with access tokens. Similarly, the second node talks about 'protection levels for cookies and security strategies for failed login attempts,' which does not relate to access token-based user authentication.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about the LLM-Anwendung (BETA)-Erweiterung or the specific steps for enabling and configuring JWT token-based authentication.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_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 answer partially addresses the question about authenticating users with access tokens, but the repeated irrelevant URL provided does not directly address the specific query.

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. For example, the first node discusses 'properties of a collection object and methods for managing collections,' which is unrelated to user authentication with access tokens. Similarly, the second node describes 'various claims that can be requested by an access token at the UserInfo endpoint,' which does not address the process of authenticating users with access tokens.

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

CustomContextualRelevancy: The retrieval context does not contain any information related to token-based authentication, JWT, or the specific steps and requirements 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: 0.8231372549019609

Individual_scores:

AnswerRelevancy: 0.8823529411764706

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.88 because the answer is mostly relevant and addresses the issue with the 'Bell' icon in XWiki, but it includes some general comments and invitations for more information 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 0.33 because the relevant node (rank 3) mentions the `notifications.enabled` setting in `xwiki.properties`, which is directly relevant. However, irrelevant nodes ranked higher (ranks 1, 2, 4, and 5) discuss topics like notification RSS feeds, overriding templates, the 'Clear All' link, and the notifications menu, which do not address enabling notifications in the `xwiki.properties` file.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports the expected output. Great job!

CustomContextualRelevancy: The retrieval context mentions the `notifications.enabled` setting in `xwiki.properties` and its default value, which aligns with the expected output. However, it does not explicitly state that this setting should be set to `true` to enable notifications.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.5650000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.375

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and there are no irrelevant statements.

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 (rank 4) that mentions the 'notifications.enabled' setting in 'xwiki.properties' is ranked lower than several irrelevant nodes. For example, the first node discusses the visual aspects of notifications, which does not address enabling notifications. Additionally, the second node lists dependencies for the notifications extension, which is unrelated to the process of enabling notifications.

ContextualRecall: The score is 0.38 because while the retrieval context supports the mention of enabling notifications in 'xwiki.properties' and the alert menu represented by a bell, it lacks information on subscribing to pages or users, watching users, receiving notifications about user events, and the 'network' tab in the user profile.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties and the alert menu, but lacks details on subscribing to pages or users and viewing the network tab.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.44000000000000006

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is completely relevant and addresses the question without any irrelevant statements. 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 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses general notification settings, the second node lists various notification customization options, the third node is about email feature customization, the fourth node describes different ways to receive notifications, and the fifth node is about overriding default notification templates. None of these nodes address enabling notifications for one's own actions.

ContextualRecall: The score is 0.00 because the retrieval context does not mention that users do not receive notifications about their own activity by default (sentence 1) or the ability to disable this filter in the 'Advanced filtering options' section of the user's notification settings (sentence 2).

CustomContextualRelevancy: The retrieval context mentions notification settings and types but lacks specific details about disabling the filter for one's own activity and advanced filtering options.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6933333333333334

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8333333333333334

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

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

Faithfulness: The score is 0.83 because the actual output incorrectly claims that XWiki allows users to subscribe to notifications on various events by default, whereas the retrieval context specifies that notifications are only sent by default when users are mentioned or when actions are performed on pages they created or updated.

ContextualPrecision: The score is 0.33 because the relevant node, which mentions 'Watch and custom filters' and is directly related to managing notification settings, is ranked third. Irrelevant nodes discussing enabling/disabling the notification module at the server level (first node), getting an RSS feed of notifications (second node), clearing and grouping notifications (fourth node), and choosing the default frequency and level of details for notification emails (fifth node) are ranked higher.

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

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_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 while the answer provides several useful links and resources related to XWiki and OpenID Connect, many of them are not directly addressing the specific configuration issue of automatically enabling user accounts upon 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. For example, the first node 'describes the OpenID Connect Authenticator extension, but it does not provide any information on configuring user accounts to be automatically enabled upon first login.' Similarly, the second node 'discusses container authentication and mentions OpenID Connect briefly, but it does not provide the specific configuration needed to enable user accounts automatically.' These irrelevant nodes should be ranked lower.

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 adjusting the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_006_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 while the answer addresses the main question about customizing user profile fields from an OpenID Connect provider, it includes several irrelevant references to the XWiki RESTful API, the Administration Application, and the Admin Guide for Authentication, which do not directly contribute to solving the specific issue.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node 'describes the OpenID Connect Authenticator extension but does not provide any information on customizing user profile fields or mapping additional fields.' Similarly, the second node 'explains various authentication methods for the XWiki RESTful API but does not mention anything about customizing user profile fields or mapping additional 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 the nodes in the retrieval context.

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

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_007_result.json

Overall_score: 0.5726923076923077
Individual_scores:
AnswerRelevancy: 0.875
Faithfulness: 1.0
ContextualPrecision: 0.25
ContextualRecall: 0.5384615384615384
CustomContextualRelevancy: 0.2
Reasons:

AnswerRelevancy: The score is 0.88 because the core instructions for configuring group synchronization are provided, but the inclusion of unrelated URLs slightly detracts from the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.25 because the relevant node with information about configuring group synchronization (rank 4) is ranked lower than several irrelevant nodes. For example, the first node describes the general capability of XWiki to authenticate using an OpenID Connect provider but does not provide specific steps or details about configuring group synchronization. Similarly, the

second node provides links to various configuration guides and examples but does not include specific instructions for configuring group synchronization.

ContextualRecall: The score is 0.54 because while several key instructions about configuring group synchronization properties (such as `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`) are supported by nodes in the retrieval context, specific procedural steps like locating and editing the `xwiki.properties` file, and saving and restarting the XWiki instance are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions group synchronization and the xwiki.properties file but lacks specific details on configuring oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping properties.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.7775757575757576

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.45454545454545453

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is mostly relevant and provides useful information, but it includes a statement about XWiki syntax that is not directly related to configuring file upload restrictions and a vague statement that lacks specific guidance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 1.00 because the relevant node explaining how to set limits on the maximum size of an attachment is ranked first, while all irrelevant nodes discussing topics like moving attachments, whitelisting mimetypes, and linking to external files are ranked lower.

ContextualRecall: The score is 0.45 because while the retrieval context covers the steps to set the maximum attachment size in the XWikiPreferences document (sentences 7-11), it does not mention configuring allowed mimetypes or editing the `xwiki.properties` configuration file (sentences 1-6).

CustomContextualRelevancy: The retrieval context mentions configuring the maximum size of an attachment and setting mimetype restrictions but lacks specific details about setting the mimetype in `xwiki.properties` and the exact byte value for 10MB.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_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 is mostly relevant and provides useful information, but the inclusion of a URL that does not directly address the specific question about denying script rights to a space administrator prevents it from being higher.

Faithfulness: The score is 1.00 because there are no contradictions. 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 'describes methods for checking access rights in a technical manner but does not address the specific question of denying script rights to a space administrator.' Similarly, the second node 'discusses the use of the checkAccess() method and its importance but does not provide information on denying script rights to a space administrator.'

ContextualRecall: The score is 0.00 because the retrieval context does not mention that certain rights cannot be denied if implied by admin rights, nor does it specify that rights implied by admin rights cannot be denied.

CustomContextualRelevancy: The retrieval context does not mention that admin rights cannot be denied or that this is not supported in XWiki.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.4618181818181819

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.91 because the answer is mostly relevant and provides useful information, but the URL provided does not directly address the configuration question about permissions in XWiki.

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

ContextualPrecision: The score is 0.20 because the relevant node in retrieval context, which provides information on how permissions are set for groups and users in XWiki, is ranked fifth. Irrelevant nodes, such as the first node explaining how users create spaces and pages, the second node discussing the parent/child relationship and nested pages, the third node listing various macros and their functions, and the fourth node explaining the link specification in XWiki syntax, are ranked 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.

CustomContextualRelevancy: The retrieval context mentions setting rights and managing spaces but lacks specific steps and details found in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_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 addresses the issue without any irrelevant statements. 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 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node is about getting a notification RSS feed and does not address enabling notifications in the wiki settings, which should be ranked lower.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications through the `notifications.enabled` setting in `xwiki.properties` or provide any information about setting `notifications.enabled` to `true` in `xwiki.properties`.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_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 while the response contains several URLs to documentation and extensions that do not directly address the issue of not receiving notifications, it still provides some relevant information related to the problem.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'various XWiki applications and extensions' which are not related to enabling notifications. The second node talks about 'access rights and executing tasks with specific user rights,' which is also not relevant. Similarly, the third node is about 'scripting macros and link renaming,' the fourth node describes 'properties of a collection in an AI application,' and the fifth node is about 'configuring permissions for guests in an open wiki,' none of which pertain to 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.

CustomContextualRelevancy: The retrieval context does not contain 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: 0.3090909090909091

Individual_scores:

AnswerRelevancy: 0.5454545454545454

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.55 because the output includes several irrelevant URLs that do not address the question about enabling notifications for one's own actions. However, the score is not lower because some relevant information is likely provided.

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. For example, the first node is about 'JWT token generation and its claims', the second node discusses 'database migration and invalid extensions', and the third node lists 'various XWiki extensions such as Office Importer and Syntax Highlighting'. These topics do not relate to notifications about user activity.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention notifications, user activity, 'Erweiterte Filteroptionen', or user notification settings in the user profile.

CustomContextualRelevancy: The retrieval context does not contain any information related to 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: 0.36875

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 provides relevant information about disabling notifications but includes multiple irrelevant URLs that do not 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. For example, the first node discusses 'Verbatim' and 'Quotations' features in XWiki Syntax, which are unrelated to notification settings. The second node describes an extension for authenticating users with JSON Web Tokens, which is not relevant to disabling notifications. Additionally, the third node lists various resources and API endpoints for managing collections and documents in the LLM Application, which do not pertain 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, 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: 0.44000000000000006

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 answer is fully relevant and directly addresses the question about configuring XWiki for automatic user activation with OpenID Connect. Great job!

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node only provides general information about the OpenID Connect Authenticator, and the second node mentions prerequisites and installation instructions but does not address the specific configuration needed to automatically activate user accounts.

ContextualRecall: The score is 0.20 because only the mention of 'xwiki.properties' in the 2nd node in the retrieval context supports the expected output, while the rest of the detailed instructions and outcomes are not covered.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.64

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and directly addresses the question without any irrelevant statements. 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node provides installation instructions and dependencies, but does not address how to customize user profile fields or map additional fields. Similarly, the second node offers general information about the OpenID Connect Authenticator without mentioning customization or mapping of user profile fields. The third node discusses container authentication, Active Directory, and LDAP, which are unrelated to the input query.

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

CustomContextualRelevancy: The retrieval context mentions 'xwiki.properties' but lacks specific details on 'oidc.user.mapping' and the example configurations provided in the expected output.

Question_language: de
Expected_answer_language: de
Answer_language: de

File: qa_017_result.json

Overall_score: 0.67
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 0.25
ContextualRecall: 1.0
CustomContextualRelevancy: 0.1
Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and addresses the question about setting up group synchronization between XWiki and an OpenID Connect provider 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 0.25 because the relevant node that mentions 'Group synchronization' and links to a section specifically about it is ranked fourth. Irrelevant nodes, such as the first node which provides general information about the OpenID Connect Authenticator, should be ranked lower as they do not address the specific steps required for group synchronization. Similarly, the second node describes general details of the OpenID Connect Authenticator without providing specific instructions, and the third node discusses other authentication methods like Active Directory and LDAP, which are not relevant to the input question.

ContextualRecall: The score is 1.00 because all sentences in the expected output are well-supported by the nodes in the retrieval context, specifically node 2, which covers all aspects of group synchronization and configuration in 'xwiki.properties'. Great job!

CustomContextualRelevancy: The retrieval context mentions group synchronization but lacks detailed steps and specific 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: 0.7468065268065269
Individual_scores:
AnswerRelevancy: 0.8333333333333334
Faithfulness: 0.8461538461538461
ContextualPrecision: 1.0
ContextualRecall: 0.45454545454545453
CustomContextualRelevancy: 0.6
Reasons:

AnswerRelevancy: The score is 0.83 because the answer provides relevant information about attachment restrictions in XWiki, but includes URLs that do not directly address the specific question about ensuring only images or PDF files with a maximum size of 10 MB can be uploaded.

Faithfulness: The score is 0.85 because the actual output incorrectly states that the maximum size for attachments is 10 MB and that only images and PDF files with a maximum size of 10 MB can be uploaded, whereas the retrieval context mentions different default maximum sizes and does not restrict

file types to only images and PDFs.

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node provides steps to set limits on the maximum size of an attachment and where the attachments will be stored, which is directly relevant to the input. The subsequent nodes discuss topics such as whitelist and blacklist configurations, managing documents, creating new pages and templates, and configuring properties of document collections, none of which are relevant to setting attachment upload restrictions or maximum sizes.

ContextualRecall: The score is 0.45 because while steps 1 through 5 in the expected output match the 1st node in the retrieval context, the introductory statements and specific configuration instructions do not have corresponding nodes in the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_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 contains multiple URLs that are unrelated to the question about denying script rights to a space administrator. However, the score is not lower because the core of the response still addresses the main 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. For instance, the first node contains 'a list of image links and references,' the second node discusses 'authentication configurations and LDAP settings,' and the third node is about 'mail templates for password reset and username retrieval.' None of these contexts address the question about denying script rights to a space administrator.

ContextualRecall: The score is 0.00 because the expected output discusses specific features of XWiki related to admin rights, which are not mentioned in any part of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to the admin rights or their denial as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.34570135746606334

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 0.8823529411764706

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.85 because the answer is mostly relevant but includes repetitive irrelevant URLs that do not directly address the specific configuration question about team access and editing permissions in XWiki.

Faithfulness: The score is 0.88 because the claim states that the Test-Benutzer or an Administrator should have no special rights for the area, but the context does not mention this restriction.

ContextualPrecision: The score is 0.00 because the nodes in the retrieval context discuss topics such as 'creating children pages and changing page access rights' (first node), 'what a wiki is and how to manage panels' (second node), 'various macros and their uses' (third node), 'applications and extensions in XWiki' (fourth node), and 'language preferences and translations in XWiki' (fifth node), none of which are relevant to configuring team-specific access control 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 specific steps for creating groups, setting wiki-wide and area-specific rights, or inheriting rights for subpages as outlined in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_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 addresses the issue about the notification bell not being available despite the extension being installed. 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 getting a notification RSS feed and settings related to notifications, but does not address enabling notifications in the wiki. The second node explains how to override default notification templates and provides installation instructions, but does not mention enabling notifications in the wiki. The third node describes the 'Clear All' link for notifications and how to manage notifications, but does not cover enabling notifications in the wiki. The fourth node is about watching locations and creating custom notification filters, but does not mention enabling notifications in the wiki. The fifth node provides a description and various settings related to notifications, but does not address enabling notifications in the wiki.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications in the wiki settings or provide instructions for setting the `notifications.enabled` property in `xwiki.properties`.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.5166666666666666

Individual_scores:

AnswerRelevancy: 0.95

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.95 because the response is highly relevant to the issue of not receiving notifications, but it includes a minor irrelevant detail about generating an access code.

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.33 because the relevant node (rank 3) mentions 'Watched Entities' and 'Follow a user,' which are pertinent to receiving notifications. However, irrelevant nodes (rank 1, 2, 4, and 5) discuss topics like RSS feeds, overriding templates, clearing notifications, and the notification menu, which should be ranked lower.

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 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: 0.4488888888888889

Individual_scores:

AnswerRelevancy: 0.4444444444444444

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.44 because the output contains multiple irrelevant statements about enabling notifications in Discord and Twitter, which do not address the user's question about enabling notifications for their own actions. However, some parts of the response may still be somewhat related to the topic of notifications.

Faithfulness: The score is 1.00 because there are no contradictions. Great job staying true to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context discuss topics such as RSS feeds, notification settings, notification preferences, overriding templates, and ways to receive notifications, none of which address enabling notifications for one's own activity. Therefore, all irrelevant nodes are ranked higher.

ContextualRecall: The score is 0.50 because the retrieval context partially aligns with the expected output, specifically regarding the filtering of one's own events, but it lacks information on disabling the filter and advanced filtering options.

CustomContextualRelevancy: The retrieval context mentions that users can filter their own events and adjust notification settings, but it does not specify that users do not receive notifications about their own activity by default or how to disable this filter in advanced settings.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7466666666666667

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.90 because the answer is mostly relevant and addresses the question about disabling notifications. However, it includes irrelevant links related to XWiki, which are not pertinent to the topic of social media notifications.

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 0.33 because the relevant node (third node) that mentions 'Auto Watch' and 'Default Watch Settings' is ranked lower than multiple irrelevant nodes. For instance, the first node discusses general notification settings but does not address how to disable notifications, and the second node provides links to notification preferences macros without explaining how to disable notifications.

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.

CustomContextualRelevancy: The retrieval context mentions auto-watched pages and notification settings but lacks details about modifying the 'Suivi automatique des pages' and removing custom filters.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_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 answer is highly relevant and addresses the main question about configuring OpenID Connect for automatic user activation. However, it includes a statement about the XWiki RESTful API, which is not directly related to the specific configuration issue.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For instance, the first node provides general information about the OpenID Connect Authenticator, such as its developer and compatibility, but does not address the specific configuration needed. Similarly, the second node discusses container authentication and mentions OpenID Connect Authentication but lacks details on automatic user activation. The third node mentions automatic synchronization but not the specific instructions needed. The fourth node discusses various authentication methods for the XWiki RESTful API without providing the required configuration details. Lastly, the fifth node lists endpoints and customization options without addressing the specific configuration needed for automatic user activation.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the `xwiki.properties` file or setting the `oidc.enableUser` property.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.5225

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 0.9375

ContextualPrecision: 0

ContextualRecall: 0.8

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the answer is mostly relevant to the question about customizing user profile fields from an OpenID Connect provider in XWiki. However, it includes URLs and information about XWiki RESTful API, XWiki Syntax, and Model Validation UI, which are not directly related to the specific query.

Faithfulness: The score is 0.94 because the actual output incorrectly states that the `OpenIDConnectProvider` extension must be installed using the Package Manager, while the retrieval context indicates it can be installed with the Extension Manager.

ContextualPrecision: The score is 0.00 because all the nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'various authentication methods for the XWiki RESTful API, including OIDC,' but does not provide information on mapping user profile fields. The second node provides details about the 'OpenID Connect Authenticator extension for XWiki,' but does not mention how to map user profile fields. The third node lists 'configuration properties related to the OpenID Connect provider and endpoints,' but does not cover user profile field mapping. The fourth node explains 'various types of references in XWiki,' which is unrelated to mapping user profile fields from an OpenID Connect provider. Finally, the fifth node discusses 'naming strategies for wiki pages in XWiki,' which is not relevant to mapping user profile fields from an OpenID Connect provider.

ContextualRecall: The score is 0.80 because most of the expected output is supported by the 3rd node in the retrieval context, including details on customizing fields, using the `oidc.user.mapping` property, providing examples, and saving and restarting XWiki. However, the retrieval context does not specifically mention replacing `myxproperty1` and `myxproperty2` with actual XWiki user property names.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.8172727272727272

Individual_scores:

AnswerRelevancy: 0.9696969696969697

Faithfulness: 0.9166666666666666

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.7

Reasons:

AnswerRelevancy: The score is 0.97 because the answer is highly relevant and provides detailed information on configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes a link to general user guide documentation, which is not directly related to the specific configuration process asked about.

Faithfulness: The score is 0.92 because the actual output incorrectly states that you need to create groups if they don't exist, while the retrieval context clarifies that groups are automatically created during synchronization.

ContextualPrecision: The score is 0.50 because the second node, which provides 'specific information about group synchronization, including the use of the `oidc.groups.claim` and `oidc.groups.mapping` properties in the `xwiki.properties` file', is correctly identified as relevant. However, the first node, which 'primarily provides links to various configuration guides and general information about OpenID Connect, but does not specifically address the steps required to configure group synchronization between XWiki and an OpenID Connect provider', is ranked higher than it should be. Additionally, the third, fourth, and fifth nodes, which provide 'general information about the OpenID Connect authenticator', 'a table of contents and links to various sections', and 'an image and a link to more information about managing users and groups in XWiki', respectively, are also ranked higher than they should be.

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

CustomContextualRelevancy: The retrieval context mentions configuring group synchronization in the xwiki.properties file and mapping groups, but lacks detailed steps and examples found in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_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 provided URLs do not directly address the question about limiting file types and sizes for uploads. However, the answer is still relatively high because it likely contains useful information related to the overall configuration and features of XWiki.

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'various macros and preferences related to notifications, office document viewing, PDF table of contents, and other unrelated topics', which does not address file upload restrictions or attachment size limits. Similarly, the second node is about 'customizing the PDF cover by overriding the template `pdfcover.vm` and adding a company logo', which is also unrelated.

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 MIME types or setting maximum upload size in XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.5733333333333334

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.67 because the answer partially addresses the question about refusing script rights to a space administrator, but includes irrelevant information such as a link to XWiki syntax documentation and a general offer for further assistance.

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

ContextualPrecision: The score is 0.00 because the nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'default right settler policies and implied rights,' which does not address the specific question. Similarly, the second node talks about 'checking access rights in the current context,' which is not specific to refusing script rights to an administrator.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports the expected output without any contradictions. Great job!

CustomContextualRelevancy: The retrieval context mentions that an admin right implies other rights and cannot be overridden at a lower level, but it does not explicitly state that admin rights cannot be refused.

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 provided URL does not directly address the question about configuring access in XWiki, leading to repeated irrelevant statements.

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. For example, the first node explains the technical details of how pages and spaces are created and managed in XWiki, but it does not provide information on setting access controls for different teams. Similarly, the second node lists various macros and their purposes in XWiki, which is not relevant to configuring access controls for team spaces.

ContextualRecall: The score is 0.00 because the retrieval context does not mention any of the specific steps or details outlined in the expected output.

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.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.5419047619047619

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.14285714285714285

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.87 because the main content is relevant and addresses the question about authenticating users with access tokens, but the repeated irrelevant URLs slightly detract 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 0.50 because the first node discusses claims via access tokens but not the authentication process, and it is ranked higher than the second node, which provides relevant information about configuring authorized applications for using access tokens to authenticate users. Additionally, nodes ranked 3, 4, and 5 provide information on a security module, security settings for cookies, and user profile visibility, none of which are relevant to the authentication process using access tokens.

ContextualRecall: The score is 0.14 because only the 2nd sentence in the expected output is partially supported by the 2nd node in the retrieval context, while the rest of the sentences lack corresponding information in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions 'Authorized Applications' and configuring properties like title, URL, and public key, but lacks details on JWT claims, enabling the authenticator, and using the token.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_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 but includes a URL that does not directly address the question about authenticating users with access tokens.

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 the nodes in the retrieval context discuss topics such as 'group synchronization', 'template customization', 'event listeners for OpenID Connect', 'protection levels for cookies', 'security strategies for failed logins', 'Kerberos SSO Authentication', 'LDAP group mapping', 'mail templates for password reset and username retrieval', and 'custom claims for OpenID Connect', which are not relevant to the question about 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 information in the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication for the LLM application (BETA) extension, JWT tokens, or the specific steps for enabling and configuring the token authenticator.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_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 provides several URLs that do not directly address the question about authenticating users with access tokens, but it still contains relevant information.

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 discusses 'properties and methods related to collections in a wiki,' which is unrelated to user authentication with access tokens. The second node talks about 'user profile claims and OAuth 2.0,' but does not address authenticating users with access tokens for the application LLM. Additionally, the third node focuses on 'default right settler policies,' which pertain to access rights and permissions, not user authentication with 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.

CustomContextualRelevancy: The retrieval context does not contain any information about authenticating users with access tokens or JWTs, which is the main focus of the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_mistral-nemo_12b_Q4

File: qa_001_result.json

Overall_score: 0.8666666666666668

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and there are no irrelevant statements.

Faithfulness: The score is 1.00 because there are no contradictions. Great job staying true to the retrieval context!

ContextualPrecision: The score is 0.33 because the relevant node, which contains information about the `notifications.enabled` setting in `xwiki.properties` (rank 3), is ranked lower than several irrelevant nodes. For instance, the first node discusses getting a notification RSS feed and settings for notifications, but it does not address enabling notifications in `xwiki.properties`. Similarly, the second node talks about overriding default notification templates and installation instructions, which are also not relevant.

ContextualRecall: The score is 1.00 because all the information needed to enable notifications in the wiki is accurately supported by the nodes in the retrieval context. Great job!

CustomContextualRelevancy: The retrieval context includes the exact setting 'notifications.enabled' in 'xwiki.properties' to 'true', which matches the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.515

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.125

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and there are no irrelevant statements in the actual output. 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 0.25 because the relevant node in the fourth rank mentions the `notifications.enabled` setting in `xwiki.properties`, which is directly relevant to enabling notifications. However, irrelevant nodes in the first, second, third, and fifth ranks discuss how notifications are displayed and managed, list dependencies, provide various ways to receive notifications, and discuss the 'What's New' application for news notifications, none of which explain how to enable notifications or ensure subscriptions.

ContextualRecall: The score is 0.12 because only the 5th node in the retrieval context mentions 'notifications.enabled' in xwiki.properties, while the rest of the expected output sentences cannot be attributed to any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and the bell icon but lacks details on subscribing to pages or users and viewing the network tab.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.44000000000000006

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and there are no irrelevant statements.

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. For example, the first node discusses 'how to get a notification RSS feed and general settings' but does not address enabling notifications for one's own actions. Similarly, the second node lists 'various customization options for notification emails' but does not mention how to enable notifications for one's own actions. The third node explains 'how to enable and disable the email feature and customize live email notifications grace time' but it does not address enabling notifications for one's own actions.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention that users do not receive notifications about their own activity unless explicitly targeted, nor do they mention the ability to disable this filter in the 'Advanced filtering options' section of the user's notification settings.

CustomContextualRelevancy: The retrieval context mentions default notification settings and the ability to change them but lacks specific details about disabling the filter for one's own activity in the Advanced filtering options.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.7333333333333333

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3666666666666667

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant with no irrelevant statements.

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

ContextualPrecision: The score is 0.37 because the relevant nodes in the retrieval context are not consistently ranked higher than the irrelevant nodes. For example, the first node discusses enabling and disabling the notification module and email notifications, which is not directly related to disabling notifications for specific pages or changes. Similarly, the second node talks about getting a notification RSS feed and the settings menu, which does not address the input query. These irrelevant nodes are ranked higher than the third and fifth nodes, which mention 'Watch and custom filters' and 'Choosing the default auto-watched mode,' both of which are directly relevant to managing and disabling notifications.

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

CustomContextualRelevancy: The retrieval context mentions settings for notifications and auto-watched mode but lacks details on major modifications, notification filters, and changing settings in the user profile.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.38666666666666666

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is mostly relevant and provides useful information, but it includes links to the XWiki RESTful API documentation and XWiki Access Rights documentation, which are not directly related to configuring automatic user account enabling upon first login using OpenID Connect.

Faithfulness: The score is 1.00 because there are no contradictions. Great job staying faithful to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'general information about the OpenID Connect Authenticator,' the second node 'mentions OpenID Connect Authentication' but lacks specific configuration details, and the third node 'discusses various authentication methods' without addressing the required 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.

CustomContextualRelevancy: The retrieval context does not contain information about adjusting the xwiki.properties file or setting the oidc.enableUser property to true.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.43508771929824563

Individual_scores:

AnswerRelevancy: 0.8421052631578947

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.84 because the answer provides useful information but includes links that are not directly relevant to the specific question about customizing user profile fields from the 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 (third node) that mentions 'the configuration of the OIDC authenticator in `xwiki.properties`' is ranked lower than irrelevant nodes. For example, the first node 'describes the general features and installation details of the OpenID Connect Authenticator but does not address customizing user profile fields', and the second node 'explains various authentication methods available in XWiki, including OpenID Connect, but does not provide information on mapping user profile fields'.

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

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.5514705882352942

Individual_scores:

AnswerRelevancy: 0.9583333333333334

Faithfulness: 0.8823529411764706

ContextualPrecision: 0.25

ContextualRecall: 0.4666666666666667

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.96 because the answer is highly relevant and provides useful information about configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes a link to the XWiki RESTful API documentation, which is not directly related to the question.

Faithfulness: The score is 0.88 because the actual output incorrectly states that group mappings can be edited in the `xwiki-authentication.properties` file, while the retrieval context specifies that group mappings are defined using the `oidc.groups.mapping` property.

ContextualPrecision: The score is 0.25 because the relevant node providing specific information about group synchronization (rank 4) is ranked lower than several irrelevant nodes. For example, the first node discusses general capabilities and installation details of the OpenID Connect Authenticator, and the second node provides links to various configuration guides, neither of which address group synchronization.

ContextualRecall: The score is 0.47 because while the retrieval context provides information on configuring group synchronization properties such as `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, it lacks details on locating and editing the `xwiki.properties` file, as well as saving and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context briefly mentions group synchronization in the xwiki.properties file but lacks detailed steps and specific properties like oidc.groups.claim,

oidc.groups.separator, and oidc.groups.mapping.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.7836363636363637

Individual_scores:

AnswerRelevancy: 0.8181818181818182

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.82 because the answer provides relevant information about file uploads but includes several irrelevant URLs and an assumption about using MediaWiki, which may not apply to the general question.

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

ContextualPrecision: The score is 1.00 because the relevant node, which mentions 'As an administrator you can set limits on the maximum size of an attachment' and provides steps to change the maximum upload size in the XWikiPreferences document, is ranked first, while all irrelevant nodes discussing topics such as moving attachments, deleting attachments, linking to files, and backlinks indexation are ranked lower.

ContextualRecall: The score is 0.50 because while the retrieval context covers the steps to set the maximum attachment size (sentences 7-12), it does not mention configuring allowed mimetypes (sentences 1-6).

CustomContextualRelevancy: The retrieval context includes steps to configure maximum attachment size and mentions mimetype restrictions, but lacks specific details on configuring allowed mimetypes in xwiki.properties and the exact value for 10MB in bytes.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.2

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because the provided output contains numerous irrelevant statements related to SharePoint and unrelated URLs, which do not address the question about denying script rights to a space administrator in XWiki.

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 'checking access rights programmatically using methods like

checkAccess and hasAccess', which does not address denying script rights to a space administrator. The second node focuses on 'the usage of checkAccess() method and its importance in security checkpoints', which is also unrelated. The third node provides 'examples of using hasAccess() and checkAccess() methods in scripts', which is not relevant. The fourth node describes 'HTTP methods and status codes for managing attachments in a wiki', which is unrelated. The fifth node describes 'HTTP methods and status codes for managing objects and their properties in a wiki', which is not relevant. The sixth node describes 'HTTP methods and status codes for managing the history of objects in a wiki', which is unrelated to the question.

ContextualRecall: The score is 0.00 because the sentence 'This is unfortunately not supported in XWiki, rights that are implied by admin right cannot be denied.' is not found in any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not mention that rights implied by admin right cannot be denied.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.44571428571428573

Individual_scores:

AnswerRelevancy: 0.9285714285714286

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.1

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is mostly relevant and provides useful information on configuring access rights in XWiki. However, it includes links about Nested Pages Migration and XWiki Syntax, which are not directly relevant to the question asked.

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 0.20 because the relevant node in retrieval context (rank 5) is ranked lower than several irrelevant nodes. For example, the first node explains how pages and spaces are created and managed in XWiki, but it does not provide information on configuring access rights for teams. Similarly, the second node discusses the parent/child relationship and the concept of nested pages in XWiki, which is not relevant to configuring access rights for teams.

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.

CustomContextualRelevancy: The retrieval context mentions setting rights and creating spaces but lacks specific steps and details found in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_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 addresses the issue without any irrelevant statements. 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 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'getting a notification RSS feed and the settings for notifications,' which does not address enabling notifications in the wiki settings. The second node talks about 'overriding default notification templates and installation instructions,' which are not relevant. The third node provides 'links to various sections about notifications,' but none mention enabling notifications. The fourth node discusses 'customizing notification emails and creating global notification filters,' which is not relevant. The fifth node describes the 'Clear All link and grouping notifications,' but does not mention enabling notifications.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications through the `xwiki.properties` file or setting `notifications.enabled` to `true`, and it lacks any information about the `xwiki.properties` file or the `notifications.enabled` setting.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3230769230769231

Individual_scores:

AnswerRelevancy: 0.6153846153846154

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.62 because although the response provides some relevant information, it includes multiple URLs that do not address the specific issue of not receiving notifications, thus reducing its overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the topic of notifications. For example, the first node discusses 'various XWiki applications like Meeting Application, Forum Application, and Ideas Application,' which are unrelated to notifications. Similarly, the second node talks about 'access rights and executing tasks with specific user rights,' which is not relevant to managing notifications. These irrelevant nodes should be ranked lower.

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 enabling notifications, following pages or users, or managing notifications in user profiles.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_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 answer partially addresses the question about enabling notifications for one's own actions, but includes multiple irrelevant URLs that do not directly relate to the issue.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'JWT token generation and its claims,' which is unrelated to notifications about user activity. Similarly, the second node is about 'upgrading wikis and handling invalid extensions,' which does not pertain to enabling notifications for user actions.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention notifications about user activity, related settings, 'Erweiterte Filteroptionen', or user profile notification settings.

CustomContextualRelevancy: The retrieval context does not contain any information related to notification settings or user activity filters.

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 actual output contains multiple irrelevant URLs that do not address the question about disabling notifications. However, the score is not lower because there is still some relevant information present.

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.00 because all the nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'XWiki syntax and verbatim content,' which is unrelated to notifications. The second node is about 'JWT authentication for the LLM Application,' which does not address notifications. The third node lists 'resources and API endpoints for the LLM Application,' which are not relevant to disabling notifications. The fourth node provides 'information for developers about building applications on XWiki,' which does not help with notifications. Lastly, the fifth node discusses 'database migration and invalid extensions,' which is not related to notifications.

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

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.48000000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and directly addresses the question without any irrelevant statements. 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 0.20 because the relevant node (rank 5) that mentions the `xwiki.properties` file is ranked lower than four irrelevant nodes. For example, the first node provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed, and the second node includes links to various guides without mentioning the `xwiki.properties` file or the `oidc.enableUser` property.

ContextualRecall: The score is 0.20 because only the mention of 'xwiki.properties' in the 2nd node in the retrieval context is relevant to the expected output, while other specific instructions and properties are not covered.

CustomContextualRelevancy: The retrieval context does not mention the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.60000000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.6

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the output is perfectly relevant and addresses the input without any irrelevant statements. 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 0.20 because the relevant node (rank 5) is ranked lower than several irrelevant nodes. For example, the first node provides 'general information about prerequisites, installation instructions, and dependencies' and should be ranked lower. Similarly, the second node 'provides a general description of the OpenID Connect Authenticator' and does not address the customization of user profile fields, so it should also be ranked lower.

ContextualRecall: The score is 0.60 because while the retrieval context supports general information about configuring the `xwiki.properties` file and using properties, it lacks explicit details on the specific

examples and instructions provided in the expected output.

CustomContextualRelevancy: The retrieval context mentions the configuration file xwiki.properties but lacks details on customizing user profile fields and the specific oidc.user.mapping property.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.5900000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and there are no irrelevant statements.

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

ContextualPrecision: The score is 0.25 because the relevant node in retrieval context (rank 4) is ranked lower than multiple irrelevant nodes. For example, the first node provides 'general information about the OpenID Connect Authenticator' and the second node describes 'general details' without addressing group synchronization, which should be ranked lower.

ContextualRecall: The score is 0.50 because while the retrieval context aligns well with the configuration steps for group synchronization (nodes 1, 3, 4, 5, 6, 7), it lacks specific instructions on locating, opening, and modifying the 'xwiki.properties' file (nodes 2, 8, 9, 10, 11, 12).

CustomContextualRelevancy: The retrieval context mentions group synchronization but lacks detailed steps and specific 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: 0.7892307692307693

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 0.8461538461538461

ContextualPrecision: 1.0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.90 because the main response is highly relevant and provides useful information on setting file upload restrictions in a wiki. However, it includes links to unrelated features, which slightly detracts from its overall relevance.

Faithfulness: The score is 0.85 because the actual output incorrectly states that only images or PDF files with a maximum size of 10 MB can be uploaded, while the retrieval context allows attachments up to 100GB by default. Additionally, the actual output mentions an automatic file size limitation feature not present in the retrieval context.

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node provides 'detailed steps on how to set limits on the maximum size of an attachment,' which is directly relevant. The irrelevant nodes, such as the second node discussing 'MIME type whitelists and blacklists,' the third node about 'managing documents,' the fourth node explaining 'how to create new pages,' and the fifth node discussing 'properties of document collections and embedding models,' are all correctly ranked lower.

ContextualRecall: The score is 0.60 because while the retrieval context provides detailed steps for configuring the maximum attachment size (steps 2.1 to 2.5), it lacks information on configuring allowed MIME types and editing the `xwiki.properties` file (step 1).

CustomContextualRelevancy: The retrieval context partially covers the expected answer. It includes steps to configure maximum attachment size and mentions MIME type restrictions but lacks specific details on configuring `xwiki.properties` and the exact MIME types allowed.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_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 actual output contains multiple irrelevant URLs that do not address the question about denying script rights to a space administrator. However, the score is not lower because the core information provided is somewhat relevant to the input.

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. For example, the first node contains a list of image links and references which are not relevant to the question about denying script rights to a domain administrator in XWiki, while the second node discusses authentication methods, specifically Kerberos and LDAP, which are also unrelated.

ContextualRecall: The score is 0.00 because the expected output discusses features and rights in XWiki that are not mentioned in any of the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information about admin rights or their denial in XWiki.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.368

Individual_scores:

AnswerRelevancy: 0.84

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.84 because the answer provides useful information but includes several irrelevant links that do not directly address the specific configuration 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 configuring access control for team-specific areas in XWiki. For example, the first node discusses additional actions like creating children pages and changing page access rights, which does not provide specific steps for the required configuration. The second node explains the concept of panels in XWiki, which is unrelated. The third node lists macros available in XWiki, which does not address the setup of access control. The fourth node talks about applications and extensions, which is not relevant. Lastly, the fifth node provides information about language support and translations, which is also unrelated.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting wiki-wide rights, setting area-specific rights, or inheriting rights for child pages.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_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 addresses the issue about the notification bell not being available despite the extension being installed. Great job!

Faithfulness: The score is 1.00 because there are no contradictions between 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 getting a notification RSS feed and settings for notifications, but does not mention enabling notifications in the `xwiki.properties` file. Similarly, the second node is about overriding default notification templates and installation instructions, which is not relevant. The third node describes the 'Clear All' link and default behavior for receiving notifications, but does not address enabling notifications in the `xwiki.properties` file. The fourth node is about custom filters and watching locations, which does not help in enabling notifications in the `xwiki.properties` file. Lastly, the fifth node lists various features and settings of the Notifications Application, but does not mention enabling notifications in the `xwiki.properties` file.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about ensuring notifications are activated in the wiki or provide information on defining the property `notifications.enabled` in `xwiki.properties`.

CustomContextualRelevancy: The retrieval context does not mention enabling notifications by 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: 0.6116666666666666

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.625

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.80 because the URL provided is not directly relevant to the issue of not receiving notifications.

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.33 because the relevant node, which mentions 'Watched Entities' and 'Follow a user,' is ranked third. Irrelevant nodes discussing 'RSS feed and settings' (first node) and 'overriding default notification templates' (second node) are ranked higher.

ContextualRecall: The score is 0.62 because while the retrieval context covers aspects of following pages, users, and using the alert menu, it lacks details on enabling notifications via `xwiki.properties` and viewing followed users in the 'network' tab.

CustomContextualRelevancy: The retrieval context mentions enabling notifications and following users, but lacks details on setting `notifications.enabled` to `true`, following pages, and the 'network' tab in the user profile.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.7533333333333333

Individual_scores:

AnswerRelevancy: 0.8666666666666667

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 1.0

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.87 because the main content is relevant and addresses the question about enabling notifications for one's own actions, but it includes links that are not directly relevant to the specific query.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect faithfulness!

ContextualPrecision: The score is 0.50 because the second node, which is relevant as it mentions 'Filter your own events,' is ranked lower than the first node, which discusses general notification settings and does not address receiving notifications for one's own activity. Additionally, the third node, which lists various notification preferences macros but does not specifically address notifications for one's own activity, is ranked higher than the relevant node.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the nodes in the retrieval context, with no unsupportive reasons.

CustomContextualRelevancy: The retrieval context mentions that users only get notifications when they are mentioned or when actions are performed on pages they created or updated, but it does not specify that users do not receive notifications about their own activity by default or how to disable this filter in

the settings.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.6257575757575757

Individual_scores:

AnswerRelevancy: 0.5454545454545454

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.75

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.55 because while the answer includes some relevant information about managing notifications, it contains multiple irrelevant URLs and statements about friend lists, which do not directly address the question about disabling notifications for pages.

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, which mentions 'Auto Watch' and 'Default Watch Settings,' is ranked third. Irrelevant nodes discussing 'getting a notification RSS feed' (rank 1), 'various notification-related macros and preferences' (rank 2), 'ways to receive notifications' (rank 4), and 'overriding default notification templates' (rank 5) are ranked higher.

ContextualRecall: The score is 0.75 because while the retrieval context nodes mention default notification settings and the ability to modify them, they do not explicitly state that users will be notified of changes on their work.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.3875

Individual_scores:

AnswerRelevancy: 0.9375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is mostly relevant and provides useful information about configuring user account activation with OpenID Connect. However, it includes a statement about the XWiki RESTful API, which is not directly related to the question asked.

Faithfulness: The score is 1.00 because there are no contradictions between 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 provides 'general information about the OpenID Connect Authenticator' and does not address the specific configuration needed. Similarly, the second node discusses 'container authentication' without providing the necessary configuration details. The third node mentions

'automatic synchronization' but lacks specific configuration information. The fourth node explains 'different authentication methods' unrelated to the required configuration. Lastly, the fifth node lists 'various endpoints and customization options' but does not mention the specific configuration needed.

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 adjusting the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.4333333333333334

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the answer provides some relevant information about customizing user profile fields from an OpenID Connect provider in XWiki, but it includes multiple URLs that are not directly related to the question, which reduces the overall relevancy.

Faithfulness: The score is 1.00 because there are no contradictions. Great job staying true to the retrieval context!

ContextualPrecision: The score is 0.33 because the relevant node in the third position mentions the `xwiki.properties` file and includes properties related to OpenID Connect, which is relevant to customizing user profile fields. However, irrelevant nodes in the first, second, fourth, and fifth positions discuss topics such as authentication methods for the XWiki RESTful API, general information about the OpenID Connect authenticator, different types of references in XWiki, and naming strategies for pages in XWiki, which should be ranked lower.

ContextualRecall: The score is 0.00 because the retrieval context does not mention any of the key points related to customizing user profile fields, mapping them in the `xwiki.properties` file, or the specific properties and steps involved.

CustomContextualRelevancy: The retrieval context does not contain information about customizing user profile fields or 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: 0.711903651903652

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 0.8888888888888888

ContextualPrecision: 0.5

ContextualRecall: 0.46153846153846156

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.91 because the answer is mostly relevant to configuring group synchronization between XWiki and an OpenID Connect provider, but it includes a statement about creating new users in XWiki, which is not directly relevant to the question.

Faithfulness: The score is 0.89 because the actual output incorrectly claims that XWiki can synchronize groups at regular intervals, while the retrieval context only mentions synchronization during authentication.

ContextualPrecision: The score is 0.50 because the second node, which addresses group synchronization by mentioning the xwiki.properties file and the oidc.groups.claim property, is correctly ranked higher. However, irrelevant nodes like the first node, which provides general information about prerequisites, installation instructions, and dependencies, and the third node, which provides general information about the OpenID Connect Authenticator, are ranked too high.

ContextualRecall: The score is 0.46 because while the 2nd node in the retrieval context provides relevant information about configuring properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, it lacks details on locating and editing the `xwiki.properties` file, replacing placeholder values, and saving and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context contains information about configuring group synchronization in the xwiki.properties file, including properties like oidc.groups.claim and oidc.groups.mapping. However, it lacks specific details on setting oidc.groups.separator and saving/restarting the instance.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3538461538461538

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.77 because the answer partially addresses the input question but includes irrelevant information about a URL that does not directly address the requirement for file type and size restrictions.

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 configuring file upload restrictions. For example, the first node discusses 'various macros and their functionalities,' which is not related to the input query. Similarly, the second node talks about 'customizing the PDF cover,' which is also unrelated.

ContextualRecall: The score is 0.00 because the retrieval context does not provide any relevant information or steps related to configuring MIME types or setting a maximum attachment size as described in the expected output.

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: 0.7666666666666667
Individual_scores:
AnswerRelevancy: 0.8333333333333334
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 1.0
CustomContextualRelevancy: 1.0
Reasons:

AnswerRelevancy: The score is 0.83 because the answer mostly addresses how to refuse script rights to a space administrator, but includes URLs that are not directly relevant to the question, which slightly detracts from its overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'default right settler additional policies', the second node explains 'usage of the `hasAccess` method and other API details', the third node provides 'a list of different rights and their descriptions', and the fourth node discusses 'link resolution and parameters'. None of these contexts address the specific question about refusing script rights to an administrator.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the information in the retrieval context, with no contradictions.

CustomContextualRelevancy: The retrieval context explicitly states that rights implied by an admin right cannot be overridden by a denial at a lower level, matching the expected answer that admin rights cannot be refused.

Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_030_result.json

Overall_score: 0.27361344537815124
Individual_scores:
AnswerRelevancy: 0.8823529411764706
Faithfulness: 0.2857142857142857
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.2
Reasons:

AnswerRelevancy: The score is 0.88 because the answer is mostly relevant and addresses the main question about configuring spaces and permissions in XWiki. However, it includes a URL that does not directly address the specific question, which slightly reduces the relevancy score.

Faithfulness: The score is 0.29 because the actual output repeatedly references 'spaces' and their management, while the retrieval context indicates that the notion of space has been removed from the UI and spaces are emulated temporarily.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node 'explains the creation of spaces and sub-pages in XWiki but does not address access control or permissions,' and the second node 'lists various macros and their purposes, which is unrelated to setting up access control for spaces 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 mentions creating spaces and managing pages but lacks specific steps for setting access rights and permissions as detailed in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.5337888198757764

Individual_scores:

AnswerRelevancy: 0.8260869565217391

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.14285714285714285

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.83 because the main content is relevant and addresses the question about authenticating users with access tokens, but the repeated irrelevant URLs lower the score.

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

ContextualPrecision: The score is 0.50 because the first node discusses claims via the UserInfo endpoint but does not address user authentication with access tokens, and the third node provides general information about permissions and security, which is not directly relevant. However, the second node provides relevant information about configuring authorized applications for authentication, which is why the score is not lower.

ContextualRecall: The score is 0.14 because only the 2nd sentence in the expected output can be linked to the 2nd node in the retrieval context, while the rest of the expected output is not supported by the retrieval context.

CustomContextualRelevancy: The retrieval context mentions authorized applications and JWT tokens but lacks details on enabling the token authenticator, specific claims, and usage instructions.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_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 answer is completely relevant and directly addresses the question about authenticating users with access tokens. Great job!

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'group synchronization and customization for OpenID Connect,' which is not related to authenticating users with access tokens. Similarly, the second node is about 'protection levels for cookies and security strategies for failed logins,' which does not address the use of access tokens for user 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 the LLM-Anwendung (BETA)-Erweiterung, JWT tokens, or the specific configuration steps mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.33999999999999997

Individual_scores:

AnswerRelevancy: 0.7

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.70 because while the answer provides some relevant information, it includes URLs that do not directly address the question about authenticating users with access tokens.

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. For example, the first node 'describes properties and methods related to a collection resource,' which does not relate to authenticating users with access tokens. Similarly, the second node 'discusses various claims that can be requested at the UserInfo endpoint by an access token,' but does not provide information on user authentication with access tokens.

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

CustomContextualRelevancy: The retrieval context does not contain any information about authenticating users with access tokens or JWTs, nor does it mention the specific steps or configurations required for this process.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Model: AI.Models.qa_mistral2_large

File: qa_001_result.json

Overall_score: 0.8666666666666668

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and addresses the question 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 0.33 because the relevant node, which mentions 'Indicates if the notification module should be enabled on the platform,' and the 'notifications.enabled' setting in 'xwiki.properties', is ranked third. Irrelevant nodes, such as the first node discussing RSS feed of notifications and settings, and the second node explaining how to override default notification templates and install extensions, should be ranked lower.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports all the sentences in the expected output. Great job!

CustomContextualRelevancy: The retrieval context includes the exact information about enabling notifications by setting 'notifications.enabled' to 'true' in 'xwiki.properties'.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.5317582417582418

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9230769230769231

ContextualPrecision: 0.25

ContextualRecall: 0.2857142857142857

CustomContextualRelevancy: 0.2

Reasons:

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

Faithfulness: The score is 0.92 because the actual output incorrectly claims you can change the default setting by switching off the 'Own event filter' in the notification settings, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.25 because the relevant node in the fourth rank mentions the setting 'notifications.enabled' in 'xwiki.properties', which directly addresses the need to ensure notifications are enabled. However, the irrelevant nodes ranked higher discuss topics like 'visual representation and grouping of notifications' (first node), 'dependencies for the notifications extension' (second node), and 'various ways to receive notifications without specific steps to enable or troubleshoot them' (third node), which do not help in resolving the issue of not receiving notifications.

ContextualRecall: The score is 0.29 because only the information about enabling notifications in xwiki.properties is supported by the retrieval context, while the rest of the details about subscribing to pages or users, watching pages or users, and the 'alert' menu are not mentioned in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties and the alert menu but lacks details on subscribing to pages or users and viewing watched users in the network tab.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.5488888888888889

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and provides useful information, but it includes a general statement about notifications in XWiki that does not directly address how to enable notifications for one's own actions.

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. For example, the first node discusses how to get a notification RSS feed and general notification settings but does not address enabling notifications for one's own actions. Similarly, the second node lists various customization options for notification emails but does not mention how to enable notifications for one's own actions.

ContextualRecall: The score is 0.50 because the retrieval context supports the information about disabling notification filters in the 'Advanced filtering options' section (sentence 2), but it does not support the information about not receiving notifications about one's own activity unless explicitly targeted (sentence 1).

CustomContextualRelevancy: The retrieval context mentions notification settings and the ability to enable/disable notifications, but it does not specifically address disabling the filter for one's own activity or the 'Advanced filtering options' section.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5633333333333334

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3666666666666667

ContextualRecall: 0.25

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and directly addresses the question about disabling notifications in XWiki.

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

ContextualPrecision: The score is 0.37 because the relevant nodes in the retrieval context, such as the third node mentioning 'Watch and custom filters' and 'Enabling notifications for all users by default,' and the fifth node discussing 'Choosing the default auto-watched mode,' are ranked lower than some irrelevant nodes. Irrelevant nodes like the first node, which talks about enabling/disabling the notification module and email notifications, and the second node, which discusses RSS feeds and general notification settings, should be ranked lower.

ContextualRecall: The score is 0.25 because only the first sentence in the expected output is somewhat supported by the 4th node in the retrieval context, while the rest of the sentences are not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions automatic notifications for actions performed by users and settings for notifications, but lacks details about major modifications, custom filters, and user profile settings.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_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 majority of the answer is relevant to configuring user accounts with OpenID Connect, but it includes some irrelevant links about the XWiki RESTful API and access rights, which slightly detract from the overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node 'only describes the OpenID Connect Authenticator extension and its installation details' and does not provide information on configuring user accounts to be automatically enabled upon first login. Similarly, the second node 'discusses container authentication and mentions OpenID Connect' but does not provide the specific configuration needed. The third node 'explains various authentication methods for the XWiki RESTful API' but does not address the required configuration. The fourth node 'is about setting access rights for terminal pages' and the fifth node 'provides information on logging in and recovering usernames and passwords', both of which are unrelated to the input query.

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

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.4346153846153847

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 0.9166666666666666

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and provides useful information on customizing user profile fields from the OpenID Connect provider in XWiki. However, it includes links to the XWiki RESTful API documentation and the Administration Application extension, which are not directly related to the specific question asked.

Faithfulness: The score is 0.92 because the actual output incorrectly claims that starting from XWiki 1.30, you can configure the OIDC authenticator directly in the wiki, while the retrieval context states that XWiki 14.10+ is required for compatibility with the OpenID Connect Authenticator.

ContextualPrecision: The score is 0.33 because the relevant node (rank 3) is ranked lower than multiple irrelevant nodes. For example, the first node discusses 'general features and installation details' which is not related to customizing user profile fields. Similarly, the second node explains 'different authentication methods' without mentioning user profile customization. The fourth node focuses on 'user and group management' and the fifth node again discusses 'various authentication methods' without relevant details. These irrelevant nodes should be ranked lower than the relevant node that mentions 'configuration of the OIDC authenticator' and 'use of configuration keys'.

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 customizing user profile fields or mapping them in the `xwiki.properties` file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.5975757575757576

Individual_scores:

AnswerRelevancy: 0.9545454545454546

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.5833333333333334

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.95 because the answer is highly relevant and provides a clear process for configuring group synchronization between XWiki and an OpenID Connect provider. However, the mention of the XWiki RESTful API is slightly off-topic and not directly related to the specific configuration process asked about.

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

ContextualPrecision: The score is 0.25 because the relevant node in retrieval context (rank 4) provides crucial information about group synchronization, including the `oidc.userinfoclaims` and `oidc.groups.mapping` properties. However, irrelevant nodes ranked higher (ranks 1, 2, and 3) contain general information about the OpenID Connect Authenticator, authentication methods, and automatic synchronization without specific details on group synchronization.

ContextualRecall: The score is 0.58 because while the retrieval context covers the main configuration steps and properties for group synchronization, it lacks specific details such as opening the file with a text editor, providing examples for replacing placeholders with actual values, and saving and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions group synchronization and the xwiki.properties file but lacks detailed steps and specific properties like oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.7823529411764706

Individual_scores:

AnswerRelevancy: 0.9117647058823529

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.91 because the answer is mostly relevant and addresses the question about limiting attachments to images or PDFs of a maximum size of 10MB. However, it includes URLs that are not directly relevant to the specific question, which slightly reduces the score.

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

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node provides detailed steps to change the maximum size of an attachment in the XWikiPreferences document, which is crucial for configuring the maximum attachment size. The subsequent nodes discuss topics such as moving and renaming attachments (second node), whitelists and blacklists for displaying attachments inline (third node), linking to external files or images (fourth node), and various ways to link to attachments, anchors, and pages (fifth node), none of which are directly relevant to setting allowed mimetypes or maximum attachment sizes.

ContextualRecall: The score is 0.50 because while the retrieval context (node 1) provides detailed steps for setting the maximum attachment size and saving the changes (Step 2, sentences 1-5), it lacks information about configuring allowed mimetypes and editing the `xwiki.properties` file (Step 1, sentences 1-3).

CustomContextualRelevancy: The retrieval context includes steps to configure maximum attachment size and mentions mimetype restrictions, but it lacks specific details about setting the mimetype restrictions in the xwiki.properties file and the exact value for 10MB in bytes.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_result.json

Overall_score: 0.39894736842105266

Individual_scores:

AnswerRelevancy: 0.8947368421052632

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.1

Reasons:

AnswerRelevancy: The score is 0.89 because the main content effectively addresses how to deny script rights to a space administrator. However, the inclusion of unrelated URLs 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For instance, the first node discusses 'methods for checking access rights and handling access denied exceptions,' which does not address the specific question about denying script rights to a space administrator. Similarly, the second node talks about 'the use of the checkAccess() method and its logging capabilities,' which is also not relevant to the input query.

ContextualRecall: The score is 0.00 because the retrieval context does not support any part of the expected output.

CustomContextualRelevancy: The retrieval context does not mention that rights implied by admin right cannot be denied.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_010_result.json

Overall_score: 0.45509578544061313

Individual_scores:

AnswerRelevancy: 0.9310344827586207

Faithfulness: 0.9444444444444444

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is mostly relevant and provides useful information on configuring access rights in XWiki, but it includes a URL that does not directly address the input question.

Faithfulness: The score is 0.94 because there is no 'Administer Space' section mentioned in the retrieval context.

ContextualPrecision: The score is 0.20 because the relevant node discussing 'how permissions are set for groups and users in XWiki' is ranked fifth, while the top four nodes are irrelevant, discussing topics like 'spaces and sub-pages', 'parent/child relationship', 'various macros', and 'link specification', which should be ranked lower.

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 mentions setting rights and creating spaces but lacks specific steps for creating groups, setting wiki-wide rights, and inheriting rights for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_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 there are no irrelevant statements.

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. For example, the first node discusses how to get a notification RSS feed and the settings for notifications, but it does not address enabling notifications in the wiki or the specific setting `notifications.enabled`. Similarly, the second node provides information on overriding default notification templates and installation instructions, but it does not mention enabling notifications in the wiki or the `notifications.enabled` setting.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications through the `notifications.enabled` setting in `xwiki.properties` or provide any information about setting `notifications.enabled` to `true` in `xwiki.properties`.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_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 answer partially addresses the issue of not receiving notifications, but the repeated irrelevant URLs lower its relevancy.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are about unrelated topics such as 'various applications in XWiki', 'access rights and executing tasks', 'link renaming in macros', 'properties of a collection in an AI application', and 'configuring permissions for guests'. None of these nodes pertain to notifications, which is the main focus of the input.

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

CustomContextualRelevancy: The retrieval context does not contain any information about enabling notifications, following pages or users, or configuring 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: 0.3411764705882353

Individual_scores:

AnswerRelevancy: 0.7058823529411765

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.71 because while the response contains some relevant information, it includes multiple URLs that do not directly address the specific issue of enabling notifications for one's own actions in XWiki.

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 about unrelated topics such as 'token generation and JWT claims' (first node), 'database migration and invalid extensions' (second node), 'XWiki extensions and their functionalities' (third node), 'LDAP group mapping and authentication' (fourth node), and 'reference forms and questionnaires in XWiki' (fifth node). These should be ranked lower than relevant nodes.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention anything about notifications, user activity, advanced filter options, or user profile notification settings.

CustomContextualRelevancy: The retrieval context does not contain any information related to 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: 0.3565217391304348

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 answer partially addresses the question about disabling notifications, but the repeated irrelevant link significantly detracts from its overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node is about 'XWiki Syntax and formatting,' which does not relate to disabling notifications. Similarly, the second node discusses 'LLM Application and JWT authentication,' which is also unrelated.

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

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.49000000000000005

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

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

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.25 because the relevant node in retrieval context (rank 4) that mentions the 'xwiki.properties' configuration file is ranked lower than irrelevant nodes. The first node describes the OpenID Connect Authenticator extension but does not address the specific configuration needed. The second node lists various sections and guides without providing the required configuration. The third node mentions other authentication methods but not the specific configuration for automatic activation of user accounts.

ContextualRecall: The score is 0.20 because while the 1st node in the retrieval context mentions 'xwiki.properties', it lacks details on adding or updating properties, the specific property `oidc.enableUser`, steps to save the file and restart XWiki, and the automatic activation of new user accounts upon first login via OpenID Connect.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.64

Individual_scores:

AnswerRelevancy: 0.8

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.80 because while the response contains some relevant information, it also includes several vague statements and headings that do not provide specific guidance relevant to the input.

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

ContextualPrecision: The score is 0.20 because the relevant node, which mentions 'Configuration files' and specifically 'xwiki.properties', is ranked fifth. Irrelevant nodes that provide general information about prerequisites, installation instructions, dependencies, compatibility, and container authentication are ranked higher, which lowers the score.

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

CustomContextualRelevancy: The retrieval context mentions the xwiki.properties file but lacks details on customizing user profile fields and the specific property oidc.user.mapping.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.6438461538461538

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.7692307692307693

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and there are no irrelevant statements in the actual output. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.25 because the relevant node, which mentions 'Group synchronization' and provides a link to more detailed instructions, is ranked fourth. Irrelevant nodes, such as the first node that primarily provides links to prerequisites, installation instructions, and dependencies, the second node that describes the OpenID Connect Authenticator extension and its installation details, and the third node that discusses container authentication and other methods like Active Directory and LDAP, are ranked higher. Additionally, the fifth node lists general configuration guides and files but does not address group synchronization specifically.

ContextualRecall: The score is 0.77 because while the retrieval context covers the configuration of group synchronization properties in detail, it lacks explicit mentions of locating, opening, and saving the `xwiki.properties` file, as well as restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions group synchronization but lacks detailed steps and specific 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: 0.7613756613756613

Individual_scores:

AnswerRelevancy: 0.9259259259259259

Faithfulness: 0.7142857142857143

ContextualPrecision: 1.0

ContextualRecall: 0.6666666666666666

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is highly relevant and addresses the question about configuring attachment settings in XWiki, but it includes links to unrelated extensions and user guides.

Faithfulness: The score is 0.71 because the actual output incorrectly states that the maximum size for attachments is 10 MB, while the retrieval context indicates that it is 100GB by default and 32MB for XWiki versions less than 10.9RC1.

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes, ensuring the most pertinent information is prioritized.

ContextualRecall: The score is 0.67 because while the retrieval context provides relevant information about editing `xwiki.properties` and configuring MIME types (2nd node), as well as steps to change the maximum attachment size (1st node), it lacks specific details such as the exact line for MIME type configuration and a comprehensive step-by-step guide for setting the maximum attachment size.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_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 answer is mostly relevant to the input question, but the repeated URL provided does not directly address the specific query about denying script rights to a space administrator.

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. For example, the first node is about 'images and links related to various projects', the second node discusses 'authentication methods and configurations', and the third node is about 'mail templates used for password reset and username recovery'. None of these contexts address the question about denying script rights to a domain administrator.

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 contain any information related to the admin rights or their limitations in XWiki as mentioned in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.3652173913043478

Individual_scores:

AnswerRelevancy: 0.8260869565217391

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the provided answer is mostly relevant but includes repeated irrelevant statements about a URL that does not directly address the configuration question about setting visibility and editing permissions in XWiki.

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. For example, the first node discusses 'additional actions that can be performed in XWiki,' which does not provide specific steps for configuring access control. Similarly, the second node talks about 'various macros and their functionalities,' which does not address the specific question of configuring access control for teams and areas.

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 wiki-wide or section-specific rights, or inheriting rights for child pages.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_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 there are no irrelevant statements.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'getting a notification RSS feed and settings for notifications', which is not directly relevant to enabling notifications in the wiki. The second node talks about 'overriding default notification templates and installation instructions', which does not address enabling notifications in the wiki. The third node describes the 'Clear All link and default notification settings', but does not provide information on enabling notifications in the wiki. The fourth node explains 'watching locations and custom filters for notifications', which does not help in enabling notifications in the wiki. The fifth node provides a 'description and various settings related to notifications', but does not mention enabling notifications in the wiki.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about ensuring notifications are activated in the wiki or how to activate them using `xwiki.properties`.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.5066666666666666

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.0

CustomContextualRelevancy: 0.3

Reasons:

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

Faithfulness: The score is 0.90 because the actual output incorrectly claims that users do not receive notifications on their wiki, while the context indicates that users do receive notifications under certain conditions.

ContextualPrecision: The score is 0.33 because the relevant node, which mentions 'Watched Entities' and 'Follow a user', is ranked third. Irrelevant nodes discussing 'getting a notification RSS feed and the settings for notifications' (first node) and 'overriding default notification templates and installation instructions' (second node) should be ranked lower.

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

File: qa_023_result.json

Overall_score: 0.7466666666666667
Individual_scores:
AnswerRelevancy: 0.9333333333333333
Faithfulness: 1.0
ContextualPrecision: 0.5
ContextualRecall: 1.0
CustomContextualRelevancy: 0.3
Reasons:

AnswerRelevancy: The score is 0.93 because the answer is mostly relevant and helpful, but it includes a link about XWiki Syntax, which is not directly related to enabling notifications for one's own actions.
Faithfulness: The score is 1.00 because there are no contradictions. Great job staying true to the retrieval context!
ContextualPrecision: The score is 0.50 because the second node, which mentions 'Filter your own events' and 'Filters', is directly relevant to the question and is ranked higher than some irrelevant nodes. However, the first node, which discusses getting a notification RSS feed and does not address notifications for one's own activity, should be ranked lower.
ContextualRecall: The score is 1.00 because the expected output is fully supported by the nodes in the retrieval context. Great job!
CustomContextualRelevancy: The retrieval context mentions default notification settings and the option to filter own events but lacks details on disabling the filter in advanced settings.
Question_language: fr
Expected_answer_language: fr
Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7531372549019608
Individual_scores:
AnswerRelevancy: 0.8823529411764706
Faithfulness: 1.0
ContextualPrecision: 0.3333333333333333
ContextualRecall: 0.75
CustomContextualRelevancy: 0.8
Reasons:

AnswerRelevancy: The score is 0.88 because the answer is mostly relevant and addresses the question about disabling notifications, but it includes URLs that are not directly related to the query.
Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect faithfulness!
ContextualPrecision: The score is 0.33 because the first two nodes discuss 'getting a notification RSS feed and the Notifications Macro' and 'various notification preferences macros', which are unrelated to disabling notifications, and should be ranked lower than the third node, which mentions 'Auto Watch'

and 'Default Watch Settings' relevant to managing automatically enabled notifications.

ContextualRecall: The score is 0.75 because the retrieval context supports the automatic addition of pages to the watch list (sentence 1 and 2) and the ability to modify notification settings (sentence 3), but it does not mention deleting existing notification filters (sentence 4).

CustomContextualRelevancy: The retrieval context mentions auto-watching pages and modifying notification settings, but lacks details on removing custom filters.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_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 answer is mostly relevant and addresses the configuration of XWiki for automatic user account activation using OpenID Connect. However, it includes some irrelevant statements such as a link to the XWiki RESTful API documentation and a configuration line that does not provide useful context.

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. For example, the first node provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed to automatically activate user accounts. Similarly, the second node discusses other authentication methods like Active Directory and LDAP, which are not relevant to the specific configuration required for OpenID Connect.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the `xwiki.properties` file or setting the `oidc.enableUser` property.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.6426666666666667

Individual_scores:

AnswerRelevancy: 0.88

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the answer is mostly relevant to the question about customizing user profile fields from OpenID Connect in XWiki, but it includes several irrelevant links that do not directly address the query.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.33 because the relevant node (third node) is ranked lower than several irrelevant nodes. For example, the first node discusses 'various authentication methods for the XWiki RESTful API' and the second node provides 'general information about the OpenID Connect Authenticator for XWiki, including installation and compatibility,' which are not related to mapping user profile fields.

ContextualRecall: The score is 1.00 because the retrieval context perfectly aligns with the expected output, covering all key points and instructions.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.7143589743589744

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.5384615384615384

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is mostly relevant and helpful, but it includes a link to a general guide on creating new users in XWiki, which is not directly related to configuring group synchronization with OpenID Connect.

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

ContextualPrecision: The score is 0.50 because the relevant node in rank 2, which mentions the default group synchronization and relevant properties in the `xwiki.properties` file, is correctly identified. However, irrelevant nodes in ranks 1, 3, 4, and 5, which discuss prerequisites, customization templates, unrelated endpoints, and general user management, should be ranked lower.

ContextualRecall: The score is 0.54 because while the retrieval context mentions configuring properties like oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping (supporting sentences 3, 5, and 7 of the expected output), it lacks explicit instructions for opening the xwiki.properties file with a text editor, replacing placeholders with actual values, and saving/restarting XWiki (unsupporting sentences 2, 4, 6, and 8).

CustomContextualRelevancy: The retrieval context mentions group synchronization and the use of the xwiki.properties file, but lacks detailed steps on configuring specific properties like oidc.groups.claim, oidc.groups.separator, and oidc.groups.mapping.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3652173913043478

Individual_scores:

AnswerRelevancy: 0.8260869565217391

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the answer partially addresses the question but includes multiple irrelevant URLs and general statements that do not provide specific steps for configuring file upload restrictions in XWiki.

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 unrelated to the input. For example, the first node is about 'various macros and functionalities in XWiki', which does not pertain to configuring file upload restrictions. Similarly, the second node discusses 'customizing the PDF cover in XWiki', which is also irrelevant.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention any of the steps or details provided in the expected output.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.635

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.88 because the answer is mostly relevant and addresses the question, but it includes repeated URLs that do not directly answer the steps to refuse script rights to a space administrator.

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. The first node discusses 'default right settler policies', the second node talks about 'checking access rights in a general context', the third node lists 'various access rights and permissions', and the fourth node discusses 'link resolution and parameters'. None of these address the specific question about denying script rights to an administrator.

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 an admin right will not be overridden by a denial at a lower level, but it does not explicitly state that admin rights cannot be refused.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.38666666666666666

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is highly relevant and provides useful information on configuring XWiki, but it includes URLs that do not directly address 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 0.00 because all nodes in the retrieval context are irrelevant to configuring access controls for team spaces in XWiki. For example, the first node discusses the creation of spaces and sub-pages in XWiki but does not address access controls, and the second node is about macros and HTML syntax, which is unrelated.

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

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.5483333333333333

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.875

ContextualPrecision: 0.5

ContextualRecall: 0.16666666666666666

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is completely relevant and directly addresses the question about authenticating users with access tokens. Great job!

Faithfulness: The score is 0.88 because the actual output incorrectly claims that the default response type can be 'id_token' or a combination like 'code,id_token', whereas the retrieval context states that the default response type is 'code' only.

ContextualPrecision: The score is 0.50 because the first node discusses claims at the UserInfo endpoint, which is not relevant to authenticating users with access tokens, and should be ranked lower. The second node is correctly ranked as it provides information about configuring 'Authorized Applications,' which is directly relevant. However, other irrelevant nodes discussing security modules, cookie protection, and user profile rights are ranked higher than they should be.

ContextualRecall: The score is 0.17 because only the sentence about configuring authorized applications (sentence 2) can be linked to the 2nd node in the retrieval context, while the rest of the sentences in the expected output are not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions authorized applications and some security settings, but lacks detailed information on JWT claims, token generation, and specific configuration steps.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_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 main content is relevant and addresses the input question, but the inclusion of headings like 'Gruppensynchronisation:', 'Anpassung:', and 'Fehlerbehebung:' detracts from the overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context discuss topics such as 'group synchronization, templates, and event listeners' (first node), 'protection levels for cookies, IP address usage, and security strategies for failed logins' (second node), 'LDAP group mapping and SSO implementation using Kerberos' (third node), 'mail templates for password reset and username retrieval' (fourth node), and 'custom claims for OpenID Connect and group synchronization' (fifth node), which are not relevant to the question about access token-based user 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 or the LLM-Anwendung (BETA)-Erweiterung.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.32756410256410257

Individual_scores:

AnswerRelevancy: 0.7916666666666666

Faithfulness: 0.8461538461538461

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.79 because while the answer partially addresses the question about authenticating users with access tokens, it includes multiple irrelevant URLs and a general summary that does not add new information.

Faithfulness: The score is 0.85 because the actual output incorrectly claims the ability to define the scope 'offline_access', which is not mentioned in the retrieval context, and it also misrepresents the handling of permissions by stating that rights are implicitly denied at the same level, contrary to 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 'properties and operations related to a collection resource,' which is unrelated to user authentication with 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.
CustomContextualRelevancy: The retrieval context does not contain any information related to authenticating users with access tokens or JWT, nor does it mention the specific steps outlined in the expected answer.
Question_language: fr
Expected_answer_language: fr
Answer_language: fr

Model: AI.Models.qa_mixtral-8x22b

File: qa_001_result.json

Overall_score: 0.8666666666666668
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 0.3333333333333333
ContextualRecall: 1.0
CustomContextualRelevancy: 1.0
Reasons:
AnswerRelevancy: The score is 1.00 because the answer is completely relevant and addresses the question without any irrelevant statements. Great job!
Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!
ContextualPrecision: The score is 0.33 because the relevant node (rank 3) that mentions the `notifications.enabled` setting in `xwiki.properties` is ranked lower than multiple irrelevant nodes. For example, the first node discusses how to get a notification RSS feed and the settings for notifications, but it does not provide information on enabling notifications in `xwiki.properties`. Similarly, the second node explains how to override default notification templates and install extensions, but it does not mention enabling notifications in `xwiki.properties`.
ContextualRecall: The score is 1.00 because the retrieval context perfectly supports the expected output. 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: en
Expected_answer_language: en
Answer_language: en

File: qa_002_result.json

Overall_score: 0.5471428571428572
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 0.25
ContextualRecall: 0.2857142857142857
CustomContextualRelevancy: 0.2
Reasons:
AnswerRelevancy: The score is 1.00 because the answer 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 that the actual output is perfectly aligned with the retrieval context. Great job!

ContextualPrecision: The score is 0.25 because the relevant node, which mentions enabling the notification module and the email feature in the `xwiki.properties` file, is ranked fourth. Irrelevant nodes, such as the first node discussing visual aspects of notifications and the second node listing dependencies for the notifications extension, are ranked higher.

ContextualRecall: The score is 0.29 because only the 2nd and 4th sentences in the expected output can be attributed to the 5th and 3rd nodes in the retrieval context respectively, while the rest of the sentences lack support from the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties and the alert menu but lacks details on subscribing to pages or users and viewing the network tab.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.44461538461538463

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.92 because the response is highly relevant and directly addresses the question about enabling notifications for one's own actions in XWiki. However, the closing statement 'If you have any further questions, please let me know!' is generic and does not contribute to the specific query.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context discuss various aspects of notifications, such as RSS feeds, customization options, email features, and notification templates, but none address enabling notifications for one's own actions.

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.

CustomContextualRelevancy: The retrieval context mentions default notification settings and the ability to enable or disable notifications but does not explicitly mention disabling the filter for one's own activity in 'Advanced filtering options'.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.5671794871794872

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 1.0

ContextualPrecision: 0.3666666666666667

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.77 because the answer provides some relevant information on how to disable notifications, but it includes several irrelevant details about default settings, appearance of the switch, and receiving notifications after enabling types, which 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 0.37 because the first two nodes discuss enabling/disabling email features and RSS feeds, which are not directly relevant to disabling notifications for specific pages. The third node, which is relevant, is ranked lower. Additionally, the fourth node talks about the 'Clear All' link and grouping notifications, which is also not relevant, while the fifth node, which is relevant, is ranked even lower.

ContextualRecall: The score is 0.50 because only some parts of the expected output are supported by the nodes in the retrieval context. Specifically, the retrieval context supports the idea of being notified about changes on your work (sentence 2) and selecting notification types (sentence 3), but it does not mention 'Automatic page watching' (sentence 1) or deleting notification filters (sentence 4).

CustomContextualRelevancy: The retrieval context mentions settings and filters for notifications but lacks specific details about automatic page watching and deleting custom filters.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_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 is mostly relevant and addresses the configuration of OpenID Connect authentication in XWiki. However, it includes links about the XWiki RESTful API and access rights, which are not directly related to the specific issue of enabling user accounts upon first login.

Faithfulness: The score is 1.00 because there are no contradictions between 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 provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed to enable user accounts upon their first login. Similarly, the second node discusses container authentication and mentions OpenID Connect briefly but does not provide the specific configuration required. These irrelevant nodes should be ranked lower than any relevant nodes.

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

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.35238095238095235

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 0.9285714285714286

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because the answer provides useful information but includes several irrelevant links that do not directly address the question about customizing user profile fields from the OpenID Connect provider.

Faithfulness: The score is 0.93 because the actual output incorrectly claims that XWiki 1.30 and later supports creating new OIDC client configurations, while the retrieval context specifies that XWiki 14.10+ is required for compatibility with the OpenID Connect Authenticator.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node 'describes the OpenID Connect Authenticator extension, its developer, and installation details, but does not mention customizing user profile fields or mapping additional fields.'

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.

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

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.47750000000000004

Individual_scores:

AnswerRelevancy: 0.9375

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and provides detailed instructions for configuring group synchronization between XWiki and an OpenID Connect provider. However, the inclusion of the URL

'<https://www.xwiki.org/xwiki/bin/view/Documentation/UserGuide/Features/XWikiRESTfulAPI>' is irrelevant 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 0.25 because the relevant node, which provides information about group synchronization including the use of the `oidc.groups.mapping` property, is ranked fourth. Irrelevant nodes, such as the first node which provides general information about the OpenID Connect Authenticator without specific steps for group synchronization, and the second node which mentions the ability to authenticate with an OpenID Connect provider but lacks specific steps for group synchronization, are ranked 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.

CustomContextualRelevancy: The retrieval context mentions group synchronization and the xwiki.properties file but lacks detailed steps and specific property configurations found in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.8088888888888889

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is highly relevant and provides useful information on setting attachment size limits and file type restrictions, but includes an unnecessary link to XWiki syntax that does not directly address 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 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node provides detailed steps on setting the maximum size of an attachment, which is directly relevant to configuring the maximum attachment size to 10MB. The subsequent nodes discuss topics such as moving and renaming attachments, whitelisting and blacklisting mimetypes, linking to external files, and linking to attachments and anchors, which are not relevant to setting upload restrictions or size limits.

ContextualRecall: The score is 0.50 because while the retrieval context provides detailed steps for configuring the maximum attachment size (Step 2, sentences 1-5), it lacks information on configuring allowed mimetypes (Step 1).

CustomContextualRelevancy: The retrieval context contains information on setting maximum attachment size and mentions mimetype restrictions but lacks specific details on configuring the `xwiki.properties` file and the exact mimetype allowList.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_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 provides useful information about denying script rights to a space administrator. However, it includes URLs that do not directly address the question, which slightly detracts from the overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect faithfulness!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses methods for checking access rights in code, which does not address the specific question about denying script rights to a space administrator. Similarly, the second node talks about the checkAccess() method and its security implications, which is also not relevant to the question.

ContextualRecall: The score is 0.00 because the retrieval context does not mention that certain rights cannot be denied if implied by admin rights, nor does it provide any information about the inability to deny rights implied by admin rights.

CustomContextualRelevancy: The retrieval context does not mention that rights implied by admin right cannot be denied.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.644

Individual_scores:

AnswerRelevancy: 0.92

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 1.0

CustomContextualRelevancy: 0.1

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and provides useful information on configuring access rights in XWiki, but it includes a URL that is not directly relevant to the specific question asked.

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

ContextualPrecision: The score is 0.20 because the relevant node discussing the configuration of wiki access rights is ranked fifth, while the first four nodes are irrelevant as they discuss topics such as the creation of spaces and pages, parent/child relationships, macros, and XWiki Syntax 2.1 links, which do not address configuring access rights for teams.

ContextualRecall: The score is 1.00 because every sentence in the expected output is supported by the retrieval context, specifically the 2nd node, which discusses access control, rights management, and setting rights at various levels in XWiki.

CustomContextualRelevancy: The retrieval context mentions setting rights and groups but lacks specific steps for creating groups, setting space-level rights, and inheriting rights for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.3762820512820513

Individual_scores:

AnswerRelevancy: 0.9583333333333334

Faithfulness: 0.9230769230769231

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.96 because the response is highly relevant and addresses the issue, but it lacks actionable information or guidance.

Faithfulness: The score is 0.92 because the actual output incorrectly states that you can use the Extension Manager to reinstall the Notifications Application, while the retrieval context specifies that installing extensions when offline is not supported.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses how to get a notification RSS feed and the settings for notifications, but it does not address the issue of the notification bell not being available or how to enable notifications in the wiki.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications via the `notifications.enabled` setting in `xwiki.properties`, nor does it provide information about setting `notifications.enabled` to `true` in `xwiki.properties`.

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

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_012_result.json

Overall_score: 0.3230769230769231

Individual_scores:

AnswerRelevancy: 0.6153846153846154

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.62 because the response contains multiple irrelevant links that do not address the issue of not receiving notifications, but it still provides some relevant information.

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. For instance, the first node discusses 'various applications in XWiki, such as Meeting Application and Forum Application,' which does not address notifications. Similarly, the second node talks about 'access rights and executing tasks with specific user rights,' which is unrelated to notifications. This pattern continues with nodes discussing topics like 'how links inside scripting macros are handled,' 'embedding model and chunking method for document collections,' and 'configuring permissions for guest users in an open wiki,' none of which are relevant to notifications.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention enabling notifications, following pages or users, or viewing user activities in the 'Network' tab.

CustomContextualRelevancy: The retrieval context does not contain any information about enabling notifications, following pages or users, or configuring notification settings as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_013_result.json

Overall_score: 0.26

Individual_scores:

AnswerRelevancy: 0.3

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.30 because the output contains multiple irrelevant URLs and statements that do not address the question about enabling notifications for one's own activity. However, there is some minimal relevance in acknowledging the issue, which prevents the score from being even 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. For example, the first node is about 'generating JWT tokens', which does not relate to notifications or user activity settings. Similarly, the second node discusses 'database migration and invalid extensions', which is also not relevant.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention notifications about user activity, events directed at the user, 'Erweiterten Filteroptionen', or user notification settings in the user profile.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_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 while the response addresses the question about disabling notifications, it includes several irrelevant URLs that do not 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. For example, the first node discusses 'XWiki Syntax, verbatim content, and quotations,' which are not related to disabling notifications. Similarly, the second node describes 'an extension for authenticating users with JSON Web Tokens,' which is also unrelated. This pattern continues with other nodes, such as the third node about 'managing collections and documents in the LLM Application,' none of which pertain to notification settings.

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

CustomContextualRelevancy: The retrieval context does not contain any information about automatic page following, notification settings, or deleting custom filters.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.48000000000000001

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and directly addresses the question 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 in the retrieval context is ranked fifth, while the first four nodes are irrelevant. The first node provides general information about the OpenID Connect Authenticator, the second node mentions various configuration guides and dependencies, the third node lists endpoints and customization options, and the fourth node discusses container authentication and other methods. These contexts do not address the specific configuration needed to automatically activate user accounts upon first login with OpenID Connect.

ContextualRecall: The score is 0.20 because only the mention of 'xwiki.properties' and configuring OpenID Connect in the 4th node in the retrieval context supports the expected output, while the rest of the specific instructions and details are not covered.

CustomContextualRelevancy: The retrieval context does not contain specific instructions about modifying the 'xwiki.properties' file or the 'oidc.enableUser' property.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.56

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.2

Reasons:

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

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to 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 provides 'general information about prerequisites, installation instructions, and dependencies' and does not address customizing user profile fields or mapping additional properties. Similarly, the second node offers a 'general description of allowing XWiki to authenticate with an OpenID Connect provider' without specific details on customizing user profile fields. The third node discusses 'container authentication and other authentication methods like Active

Directory and LDAP' but does not mention customizing user profile fields or mapping additional properties from an OpenID Connect provider. The fourth node lists 'various endpoints and customization options' but lacks specific information on customizing user profile fields or mapping additional properties. Finally, the fifth node provides a 'general overview of the OpenID Connect Authenticator' without addressing the customization of user profile fields or mapping additional properties.

ContextualRecall: The score is 0.60 because while the retrieval context mentions 'xwiki.properties' and customization (1st node), it lacks specific examples of user mapping properties and instructions for replacing placeholders.

CustomContextualRelevancy: The retrieval context mentions xwiki.properties and customization but lacks specific details on oidc.user.mapping and example configurations.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.5776923076923077

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.5384615384615384

CustomContextualRelevancy: 0.1

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully 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. Great job on maintaining perfect alignment with the retrieval context!

ContextualPrecision: The score is 0.25 because the relevant node (rank 4) that mentions 'Group synchronization' and provides a link to more detailed information is ranked lower than several irrelevant nodes. For example, the first node provides general information about OpenID Connect and its integration with XWiki but does not give specific steps or details about configuring group synchronization. Similarly, the second node describes the OpenID Connect Authenticator extension and its compatibility but does not provide instructions on setting up group synchronization.

ContextualRecall: The score is 0.54 because while the retrieval context nodes mention configuring group synchronization properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, they do not explicitly cover steps like locating and opening the `xwiki.properties` file, replacing placeholders with actual values, or saving and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions group synchronization but lacks detailed steps and specific 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: 0.693409090909091

Individual_scores:

AnswerRelevancy: 0.8125

Faithfulness: 0.7

ContextualPrecision: 1.0

ContextualRecall: 0.454545454545453

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.81 because the answer provides relevant information but includes multiple URLs that are not directly relevant to the specific instructions needed to ensure only images or PDF files with a maximum of 10 MB can be uploaded.

Faithfulness: The score is 0.70 because the actual output incorrectly adds a step about adding a whitelist configuration in the `WEB-INF/xwiki.properties` file, and it imposes restrictions on file types and sizes that are not mentioned in the retrieval context.

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. The first node provides detailed steps on setting the maximum size of an attachment, while the subsequent nodes discuss topics like MIME type whitelisting, managing documents, creating new pages, and configuring properties of document collections, which are not relevant to the input question.

ContextualRecall: The score is 0.45 because while the retrieval context provides detailed steps for configuring the maximum attachment size (related to sentences 8-13 in the expected output), it lacks information on configuring allowed MIME types and editing the `xwiki.properties` file (related to sentences 1-7 in the expected output).

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_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 but includes multiple irrelevant links that do not pertain to denying script rights to a space administrator.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to 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 are unrelated to the topic of denying script rights to a domain administrator. The second node discusses authentication configurations and LDAP settings, which do not address the question about denying script rights to a domain administrator. The third node talks about mail templates for password reset and username recovery, which are not relevant to the question about script rights. The fourth node provides installation instructions and dependencies for an extension, which do not relate to the question about script rights. The fifth node describes form utilization on XWiki.org, which is not relevant to the question about denying script rights to a domain administrator.

ContextualRecall: The score is 0.00 because the expected output sentence is not found in any part of the nodes in retrieval context, which mainly discuss image links, LDAP configuration, and form utilization.

CustomContextualRelevancy: The retrieval context does not contain any information about admin rights or their implications in XWiki.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.3529411764705882

Individual_scores:

AnswerRelevancy: 0.7647058823529411

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.76 because the answer partially addresses the configuration question, but the repeated irrelevant URLs lower its 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 0.00 because all nodes in the retrieval context are irrelevant. The first node discusses 'creating children pages, changing page access rights, and other advanced features', which does not address configuring access for teams in XWiki. The second node provides information about 'different macros and their functionalities', which is not related to team access and permissions. The third node explains 'various aspects of XWiki Syntax 2.1', but does not mention configuring access for teams. The fourth node talks about 'XWiki applications and extensions', which is unrelated to team-specific access permissions. The fifth node is about 'language preferences and translations', which does not help in configuring access control for teams.

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

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting wiki-wide rights, setting area-level rights, or inheriting rights for child pages.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_021_result.json

Overall_score: 0.38461538461538464

Individual_scores:

AnswerRelevancy: 0.9230769230769231

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.92 because the response is mostly relevant but includes a URL that does not directly address the issue of the notification bell not being available.

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. For example, the first node discusses getting a notification RSS feed, which does not address

enabling notifications in the wiki. Similarly, the second node is about overriding default notification templates and installation instructions, which are also not relevant.

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.

CustomContextualRelevancy: The retrieval context does not mention the `notifications.enabled` property or how to set it to `true` in `xwiki.properties`.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.6133333333333334

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.9

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.3333333333333333

CustomContextualRelevancy: 0.5

Reasons:

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

Faithfulness: The score is 0.90 because the actual output incorrectly states 'Vous ne recevez pas de notifications.', while the context indicates that users do receive notifications by default under certain conditions.

ContextualPrecision: The score is 0.33 because the relevant node, which mentions 'Watched Entities' and 'Follow a user', is ranked third. Irrelevant nodes, such as the first node discussing RSS feed settings and the second node about overriding notification templates, should be ranked lower.

ContextualRecall: The score is 0.33 because only a few aspects of the expected output are supported by the nodes in the retrieval context, such as enabling notifications by clicking on the bell and following a user. However, key details about 'notifications.enabled', 'xwiki.properties', specific instructions on following pages or users, and viewing user events in the 'réseau' tab are missing.

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.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.6177777777777778

Individual_scores:

AnswerRelevancy: 0.8888888888888888

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.89 because the answer is mostly relevant and helpful, but the inclusion of the URL

'<https://www.xwiki.org/xwiki/bin/view/Documentation/UserGuide/Features/XWikiSyntax/>' is not related to enabling notifications for one's own actions.

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. For example, the first node discusses getting an RSS feed of notifications and does not address receiving notifications about one's own activity. The second node provides a list of links and descriptions related to notifications but does not specifically address the issue of receiving notifications about one's own activity. The third node lists various notification preferences macros but does not provide information on how to enable notifications for one's own actions. The fourth node talks about overriding default notification templates and installation instructions, which are not relevant to the question about receiving notifications for one's own activity. The fifth node lists different ways to be notified in XWiki, but it does not specifically address how to enable notifications for one's own actions.

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

CustomContextualRelevancy: The retrieval context mentions default notification settings and the ability to filter notifications, but it lacks details about disabling the filter in 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.815897435897436

Individual_scores:

AnswerRelevancy: 0.8461538461538461

Faithfulness: 1.0

ContextualPrecision: 0.5833333333333333

ContextualRecall: 0.75

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.85 because the answer mostly addresses the question about disabling notifications, but includes irrelevant URLs that do not directly help with the issue.

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.58 because the first node talks about 'getting a notification RSS feed and the Notifications Macro,' which is not directly related to disabling notifications, and should be ranked lower. Additionally, the fourth node discusses 'various ways to receive notifications, such as through RSS feeds and mentions,' but does not provide specific instructions on how to disable them, and should also be ranked lower. The fifth node is about 'overriding default notification templates and installation instructions,' which is not relevant to disabling notifications, and should be ranked lower as well.

ContextualRecall: The score is 0.75 because the retrieval context partially supports the expected output. Specifically, nodes in the retrieval context mention default notification settings and accessing notification settings, but there is no mention of being notified specifically about changes on your work.

CustomContextualRelevancy: The retrieval context includes information about auto-watching pages and modifying notification settings, but lacks explicit mention of removing filters in the 'Filtres personnalisés' list.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.45499999999999996

Individual_scores:

AnswerRelevancy: 0.875

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.88 because the answer is mostly relevant and provides useful information about configuring OpenID Connect for automatic user activation. However, it includes an irrelevant link to the XWiki RESTful API documentation, which doesn't address 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 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node provides general information about the OpenID Connect Authenticator, but does not address the specific configuration needed to automatically activate user accounts. Similarly, the second node discusses container authentication and mentions OpenID Connect in passing, but does not provide the specific configuration details required to solve the problem.

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention 'xwiki.cfg configuration file', 'OpenId Connect Authentication', and 'automatic synchronization', they do not provide specific instructions to add or update properties, uncomment settings, or restart the XWiki instance as detailed in the expected output.

CustomContextualRelevancy: The retrieval context does not mention adjusting the 'xwiki.properties' file or the 'oidc.enableUser' property.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.5628019323671498

Individual_scores:

AnswerRelevancy: 0.8695652173913043

Faithfulness: 0.9444444444444444

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.87 because the main response addresses the customization of user profile fields from an OpenID Connect provider in XWiki, but it includes irrelevant URLs that do not pertain to the specific question asked.

Faithfulness: The score is 0.94 because the actual output incorrectly specifies 'OpenID Connect Generic Authenticator' instead of 'OpenID Connect Authenticator' as mentioned in 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 discusses 'various authentication methods for the XWiki RESTful API, including OIDC,' but does not provide information on customizing user profile fields or mapping them. Similarly, the second node provides 'general information about the OpenID Connect Authenticator for XWiki, including installation and compatibility,' but does not address the customization of user profile fields or mapping them.

ContextualRecall: The score is 1.00 because all sentences in the expected output are fully supported by the information in the 3rd node in the retrieval context, with no discrepancies.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.7257142857142858

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.5714285714285714

CustomContextualRelevancy: 0.7

Reasons:

AnswerRelevancy: The score is 0.86 because the answer is mostly relevant to configuring group synchronization between XWiki and an OpenID Connect provider, but it includes a link to a general user guide on creating new users, which is not directly relevant.

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

ContextualPrecision: The score is 0.50 because the second node, which is relevant and mentions the default group synchronization and relevant properties in the `xwiki.properties` file, is ranked lower than the first node, which primarily provides links to prerequisites, installation instructions, and other general information about the OpenID Connect Authenticator. Additionally, the third, fourth, and fifth nodes provide general information, features, and links but do not address the specific configuration steps for group synchronization, and should be ranked lower.

ContextualRecall: The score is 0.57 because while the 2nd node in the retrieval context covers key properties like `xwiki.properties`, `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, it lacks details on steps like opening the file, replacing placeholders, and restarting XWiki.

CustomContextualRelevancy: The retrieval context mentions group synchronization and configuring `oidc.groups.mapping` in `xwiki.properties`, but lacks detailed steps and examples for setting `oidc.groups.claim` and `oidc.groups.separator`.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.36470588235294116

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 answer is mostly relevant but includes repeated irrelevant URLs that do not address the specific question about limiting file types and sizes for attachments.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'system preferences and features related to notifications, office document viewing, and PDF table of contents,' which do not pertain to configuring file upload restrictions. Similarly, the second node is about 'customizing the PDF cover by overriding a template and adding a company logo,' which is also unrelated.

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 MIME types or setting maximum attachment size.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.7647058823529411

Individual_scores:

AnswerRelevancy: 0.8235294117647058

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.82 because the provided URLs do not directly address the question of how to refuse script rights to a space administrator, but the main content is still relevant and helpful.

Faithfulness: The score is 1.00 because there are no contradictions between 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 'default right settler policies', the second node is about 'executing tasks with specific user rights', the third node describes 'access checks and security checkpoints', the fourth node lists 'various types of rights and permissions', and the fifth node explains 'link resolution and parameters in XWiki'. None of these contexts address the issue of refusing script rights to an administrator.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports the expected output. Great job!

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_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 answer is mostly relevant but includes a URL that does not directly address the question about configuring permissions in XWiki.

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. For example, the first node discusses the creation of spaces and pages in XWiki but does not provide specific instructions on setting access rights for teams. The second node lists various macros and their uses in XWiki, which is unrelated to configuring team access rights. Additionally, the third node explains the parent/child relationship and the Nested Pages concept in XWiki, but it does not address how to set access rights for teams.

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.

CustomContextualRelevancy: The retrieval context does not contain information about creating groups, defining rights at the wiki and space levels, or inheriting rights for child pages.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.4677777777777778

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 0.8888888888888888

ContextualPrecision: 0.5

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.75 because the answer partially addresses the question about authenticating users with access tokens, but includes multiple irrelevant URL statements that do not directly contribute to the explanation.

Faithfulness: The score is 0.89 because the actual output claims that you can configure the scope of the access token in XWiki, but the retrieval context does not mention this configuration.

ContextualPrecision: The score is 0.50 because the relevant node in the second rank correctly describes the 'Authorized Applications' configuration, which is part of the process for using tokens to authenticate users. However, irrelevant nodes in the first, third, fourth, and fifth ranks discuss various unrelated topics such as claims at the UserInfo endpoint, the Security Module, cookie protection, and user profile visibility, which should be ranked lower.

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 mentions 'Authorized Applications' and some configuration details, but lacks specific information about JWT claims, token signing, and the detailed steps for enabling and using the token authenticator.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_032_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 to the question about authenticating users with access tokens. Great job!

Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect faithfulness!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'group synchronization and customization for OpenID Connect,' which is not relevant to authenticating users with access tokens. Similarly, the second node is about 'cookie protection levels and security strategies,' which does not relate to access token-based user authentication.

ContextualRecall: The score is 0.00 because the retrieval context does not mention any of the key concepts or instructions related to 'Benutzer mit Zugriffstoken', 'Token-basierten Authentifizierung', 'externen Anwendungen', 'JWT', or the specific steps and limitations of the authenticator.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication, JWT, or the specific steps for using the LLM-Anwendung (BETA) extension.

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 while some relevant information is provided, multiple irrelevant URL links are included that do not directly address the question about authenticating users with access tokens.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to 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 properties of a collection object and methods to manage collections, which is unrelated to user authentication with access tokens. Similarly, the second node discusses various claims that can be requested at the UserInfo endpoint by an access token, but it does not provide information on authenticating users with access tokens.

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

CustomContextualRelevancy: The retrieval context does not contain any information about authenticating users with access tokens, JWT, or the specific steps and properties 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: 0.8666666666666668

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and there are no irrelevant statements.

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 0.33 because the relevant node (rank 3) that mentions the 'notifications.enabled' setting in 'xwiki.properties' is ranked lower than irrelevant nodes. For example, the first node discusses getting a notification RSS feed and the settings for notifications, and the second node talks about overriding default notification templates and installation instructions, both of which are not relevant to enabling notifications in the 'xwiki.properties' file.

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: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.5385714285714285

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.14285714285714285

CustomContextualRelevancy: 0.3

Reasons:

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

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

ContextualPrecision: The score is 0.25 because the relevant node, which mentions the 'notifications.enabled' setting in 'xwiki.properties', is ranked fourth. Irrelevant nodes, such as the first node discussing the 'Clear All' link and grouping of notifications, and the second node listing dependencies for the notifications extension, should be ranked lower.

ContextualRecall: The score is 0.14 because only the mention of 'notifications.enabled' in xwiki.properties (sentence 2) is supported by the 5th node in the retrieval context, while all other details in the expected output lack corresponding information in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties and the alert menu, but lacks details on subscribing to pages or users and the network tab in the user profile.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_003_result.json

Overall_score: 0.4094736842105263
Individual_scores:
AnswerRelevancy: 0.9473684210526315
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.1
Reasons:

AnswerRelevancy: The score is 0.95 because the answer is highly relevant and directly addresses the question about enabling notifications for one's own actions in XWiki. However, it includes a general reminder about the potential downside of notifications, which is not directly related to the specific query.

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. For example, the first node discusses getting a notification RSS feed and general notification settings, but does not mention enabling notifications for one's own actions. Similarly, the second node lists various notification settings and customization options but does not address receiving notifications for one's own activity.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about not receiving notifications about one's own activity or the ability to disable this filter in the 'Advanced filtering options' section of the user profile.

CustomContextualRelevancy: The retrieval context mentions notification settings but does not specify the ability to disable the filter for one's own activity in the 'Advanced filtering options' section.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_004_result.json

Overall_score: 0.6577777777777778
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1.0
ContextualPrecision: 0.5888888888888889
ContextualRecall: 0.5
CustomContextualRelevancy: 0.2
Reasons:

AnswerRelevancy: The score is 1.00 because the response is fully relevant and directly addresses the question 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 0.59 because the first node discusses enabling and disabling the notification module at the server level, which is not directly related to user-specific notification settings, and should be ranked lower. Similarly, the fourth node discusses clearing notifications and grouping similar notifications, which is not directly related to changing notification settings or disabling notifications, and should also be ranked lower. However, the second, third, and fifth nodes are correctly

identified as relevant, discussing the notifications menu, settings, and managing automatic page watching and custom filters.

ContextualRecall: The score is 0.50 because while the retrieval context (node 5) supports the idea of notifications for pages users created or updated (related to sentence 1 of the expected output), it does not provide information on changing the setting for 'Automatic page watching' or deleting existing notification filters (related to sentences 3 and 4).

CustomContextualRelevancy: The retrieval context mentions settings and filters but lacks details on automatic page watching and major modifications.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_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 answer addresses the configuration of OpenID Connect authentication and automatic user enablement, but includes some irrelevant details about the XWiki RESTful API, access rights, and an unrelated property format.

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 configuring user accounts to be automatically enabled upon first login using OpenID Connect. For example, the first node describes the OpenID Connect Authenticator but does not provide any information about configuring user accounts, and the second node is about container authentication, which does not mention OpenID Connect or enabling user accounts on first login.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the xwiki.properties file or the oidc.enableUser property.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_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 provides relevant information about customizing user profile fields from an OpenID Connect provider, but it includes several URLs that are

not directly related to the specific question asked.

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. For example, the first node describes the OpenID Connect Authenticator, its developer, and installation details, but does not mention customizing user profile fields or mapping additional fields. Similarly, the second node discusses various authentication methods for the XWiki RESTful API, including OIDC, but does not provide information on customizing or mapping user profile fields.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about customizing user profile fields or mapping them in the `xwiki.properties` file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.5124358974358975

Individual_scores:

AnswerRelevancy: 0.9583333333333334

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.15384615384615385

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.96 because the provided URL for the XWiki RESTful API is not directly relevant to configuring group synchronization between XWiki and an OpenID Connect provider. However, the rest of the information is highly relevant and useful.

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 providing information about group synchronization ('the use of the `oidc.groups.mapping` property and the `xwiki.properties` file') is ranked fourth, while several irrelevant nodes ('general functionality of the OpenID Connect Authenticator', 'ability of XWiki to authenticate using OpenID Connect', 'different authentication methods', 'customization options', and 'various authentication methods for the XWiki RESTful API') are ranked higher.

ContextualRecall: The score is 0.15 because the retrieval context only partially supports the expected output, specifically in mentioning the `oidc.groups.claim` and `oidc.groups.mapping` properties (sentence 3 and 7), but lacks detailed instructions and examples provided in the expected output.

CustomContextualRelevancy: The retrieval context mentions group synchronization and the use of the xwiki.properties file but lacks detailed steps and specific properties like `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.6296989966555184

Individual_scores:

AnswerRelevancy: 0.08695652173913043

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.46153846153846156

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.09 because the majority of the statements are about MediaWiki, which is not relevant to the wiki platform in question, and do not address the specific requirement of setting upload limits for attachments. However, the low score indicates that there might be a minimal connection to the general topic of managing wiki settings.

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 are ranked higher than irrelevant nodes. The first node provides detailed steps on setting the maximum size of an attachment, which is directly relevant to the input. The subsequent nodes discuss topics such as moving attachments, deleting attachments, linking to external files, and linking to email addresses, which are not relevant to setting file type or size restrictions.

ContextualRecall: The score is 0.46 because while the retrieval context provides detailed steps for configuring the maximum attachment size (Steps 2.1 to 2.5 in the expected output), it lacks information on configuring allowed mimetypes and editing the `xwiki.properties` file (Step 1 in the expected output).

CustomContextualRelevancy: The retrieval context provides steps to change the maximum attachment size and mentions mimetype restrictions but lacks specific details on configuring the `xwiki.properties` file and setting the size to 10MB.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_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 answer is mostly relevant and helpful, but the provided URL does not directly address the specific question about denying script rights to a space administrator.

Faithfulness: The score is 1.00 because there are no contradictions present. 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 is about 'checking access rights programmatically using methods like checkAccess and hasAccess,' which does not address denying script rights to a space administrator. Similarly, the second node discusses 'the usage of checkAccess() method and its importance in security,' which is also unrelated. The third node provides 'examples of using hasAccess() and checkAccess() methods for different entities and actions,' but does not provide information about denying script rights to a space administrator. The fourth and fifth nodes describe 'HTTP methods and status codes for managing attachments and objects in a wiki,' which are completely unrelated to the question.

ContextualRecall: The score is 0.00 because the retrieval context does not provide any information about the inability to deny rights implied by admin rights in XWiki or about the rights implied by admin rights.

CustomContextualRelevancy: The retrieval context does not mention that rights implied by admin right cannot be denied.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.4466666666666666

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 0.9166666666666666

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and provides useful information, but it includes a URL that does not directly address the specific configuration question about access rights in XWiki.

Faithfulness: The score is 0.92 because the actual output incorrectly claims that RBAC permissions and restrictions are assigned at the page level in XWiki, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.20 because the relevant node discussing 'how permissions are set for groups or users' is ranked 5th, while the higher-ranked nodes discuss unrelated topics such as 'spaces and pages creation and management' (1st node), 'parent/child relationship and nested pages' (2nd node), 'various macros and their functions' (3rd node), and 'XWiki Syntax 2.1 Link Specification' (4th node).

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

CustomContextualRelevancy: The retrieval context mentions setting rights and creating spaces but lacks specific steps about creating groups, setting wiki-wide rights, and inheriting rights for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_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 answer is perfectly relevant and there are no irrelevant statements. 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 0.00 because all nodes in the retrieval context are irrelevant to the input. The first node discusses notification RSS feed and settings but does not address the notification bell issue. The second node talks about overriding notification templates and installation instructions without mentioning enabling notifications. The third node lists various notification features and settings

but lacks information on enabling notifications. The fourth node covers customization options without addressing enabling notifications. The fifth node describes the 'Clear All' link and notification display without providing information on enabling notifications.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications via the `notifications.enabled` setting in `xwiki.properties`, nor does it provide information about setting `notifications.enabled` to `true` in `xwiki.properties`.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.275

Individual_scores:

AnswerRelevancy: 0.375

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.38 because the actual output contains multiple URLs that do not directly address the issue of not receiving notifications, making the response largely irrelevant.

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. For instance, the first node discusses 'various XWiki applications and extensions,' which is not relevant to setting up notifications. Similarly, the second node talks about 'access rights and executing tasks with specific user rights,' which does not relate to enabling notifications. The third node explains 'how links inside scripting macros are handled and renamed,' which is not pertinent to the topic of setting up notifications.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about enabling notifications or following pages and users.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_result.json

Overall_score: 0.3166666666666667

Individual_scores:

AnswerRelevancy: 0.5833333333333334

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.58 because the provided URLs do not address the specific question about enabling notifications for one's own activity, making the response partially irrelevant.

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 'token generation and JWT claims', which is unrelated to notifications for user activity. The second node discusses 'database migration and invalid extensions', which is not relevant to activating notifications for user actions. The third node lists 'various XWiki extensions and their functionalities', which does not relate to enabling notifications for a user's own activity.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about notifications, user activity, filter options, or notification settings in user profiles.

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: 0.35

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 answer provides relevant information about disabling notifications, but it includes multiple irrelevant URLs that do not 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 0.00 because all the nodes in the retrieval context are irrelevant to the input. For example, the first node discusses 'XWiki syntax and formatting,' which is unrelated to disabling notifications. Similarly, the second node describes 'an extension for authenticating users with JSON Web Tokens,' which is not relevant to the question about disabling notifications.

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

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_015_result.json

Overall_score: 0.4288888888888889

Individual_scores:

AnswerRelevancy: 0.9444444444444444

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.94 because the output is highly relevant and addresses the question about configuring XWiki for automatic user activation with OpenID Connect. However, the placeholder comment " is irrelevant and slightly detracts from the overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed. The second node lists various guides and dependencies but does not mention the required configuration. The third node mentions the 'xwiki.cfg' file but not the 'xwiki.properties' file or the specific property 'oidc.enableUser'. The fourth node provides information about other authentication methods but not the specific configuration for OpenID Connect. The fifth node mentions 'xwiki.properties' but not the specific property or steps needed.

ContextualRecall: The score is 0.20 because while the 2nd node in the retrieval context mentions 'xwiki.properties' and configuration related to OpenID Connect, it lacks specific instructions, properties, and details mentioned in the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the 'xwiki.properties' file or the 'oidc.enableUser' property.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.6094117647058823

Individual_scores:

AnswerRelevancy: 0.6470588235294118

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.65 because several irrelevant statements related to 'java', '"', and other unrelated tags were included, which do not address the customization of user profile fields in XWiki from an OpenID Connect provider. However, the score is not lower because there is some relevant information that partially addresses the input question.

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

ContextualPrecision: The score is 0.20 because the relevant node (rank 5) mentions the configuration files, including 'xwiki.properties', which is pertinent to mapping user profile fields. However, irrelevant nodes (ranks 1 to 4) are ranked higher and discuss topics such as prerequisites, general functionality, container authentication, and various endpoints, which do not address the customization of user profile fields.

ContextualRecall: The score is 1.00 because all sentences in the expected output are fully supported by the information provided in the 2nd node in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions configuration files and xwiki.properties but lacks specific details on customizing user profile fields and the oidc.user.mapping property.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.6407936507936508

Individual_scores:

AnswerRelevancy: 0.8095238095238095

Faithfulness: 0.9444444444444444

ContextualPrecision: 0.25

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.81 because the main content addresses the process of setting up group synchronization between XWiki and an OpenID Connect provider, but it includes several irrelevant statements that do not directly contribute to the explanation.

Faithfulness: The score is 0.94 because the actual output incorrectly instructs to go to 'Tools' and select 'Plugin Manager' to find 'OpenID Connector', which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.25 because the relevant node mentioning 'Group synchronization' and providing a link to more details is ranked fourth. Irrelevant nodes, such as the one in the first rank that primarily provides links to prerequisites and general descriptions, should be ranked lower.

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

CustomContextualRelevancy: The retrieval context mentions group synchronization but lacks detailed steps and specific 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: 0.8154545454545454

Individual_scores:

AnswerRelevancy: 0.75

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.72727272727273

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.75 because while the answer provides some relevant information, it includes multiple irrelevant links and a partially cut-off statement that do not directly address how to restrict uploads to images or PDFs with a maximum size of 10 MB.

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 are ranked higher than irrelevant nodes. The first node explains how to set limits on the maximum size of an attachment, which is directly relevant to configuring the maximum attachment size. The subsequent nodes discuss topics like MIME type configurations, managing documents, creating new pages, and configuring properties of document collections, none of which are relevant to uploading attachments or setting size and type restrictions.

ContextualRecall: The score is 0.73 because while many sentences in the expected output match nodes in the retrieval context, some introductory and step-introducing sentences do not align with any specific nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context contains steps to change the maximum upload size and mentions MIME type restrictions, but does not specify the exact MIME types or the 10 MB limit.

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 answer contains several URLs that are irrelevant to the question about denying script rights to a domain administrator. However, the answer still provides some relevant information, which is why it is not scored lower.

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 contains only links to images and references, which are not relevant to the question about denying script rights to a domain administrator. The second node discusses authentication using Firefox and LDAP, which is also not related to the input.

ContextualRecall: The score is 0.00 because the sentence in the expected output does not match any part of the provided retrieval context, which discusses image links, LDAP configuration, and form utilization, none of which relate to XWiki rights.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.3384615384615385

Individual_scores:

AnswerRelevancy: 0.6923076923076923

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.69 because while the answer attempts to address the configuration question, it includes a URL that does not directly address the specific configuration question asked, which reduces its relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node discusses actions like creating children pages and changing page access rights, which does not specifically address configuring access control for teams in XWiki. The second node describes various macros and their functionalities, which is unrelated to setting up team-specific access controls. Additionally, the third node provides information about XWiki Syntax 2.1, which is not relevant to configuring access control for teams.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting wiki-wide or area-specific rights, or inheriting rights for subpages.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_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 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 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses how to get a notification RSS feed and settings related to notifications, but it does not address enabling notifications in the wiki settings.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about ensuring notifications are activated in the wiki or how to activate them, nor does it provide information about setting the `notifications.enabled` property in `xwiki.properties` to `true`.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.41999999999999993

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.0

CustomContextualRelevancy: 0.1

Reasons:

AnswerRelevancy: The score is 0.67 because the output includes a URL that does not directly address the issue of not receiving notifications.

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 (third node) is ranked lower than several irrelevant nodes. For example, the first node discusses 'getting a notification RSS feed and the settings for notifications,' which is not about enabling notifications or following pages/users. Similarly, the second node is about 'overriding default notification templates and installation instructions,' which is

also not relevant. The relevant node should be ranked higher to improve the 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 mentions enabling notifications and following users but lacks detailed steps on setting `notifications.enabled` in `xwiki.properties` and specific instructions on following pages or users.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.4822222222222222

Individual_scores:

AnswerRelevancy: 0.1111111111111111

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.11 because the majority of the statements provided are irrelevant to the specific question about enabling notifications for one's own actions on XWiki. The answer includes unrelated URLs and references to Google or Facebook settings, which do not address the user's query.

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 second node, which mentions 'Filter your own events,' is relevant and ranked higher than some irrelevant nodes. However, irrelevant nodes like the first node, which discusses getting a notification RSS feed and general notification settings, and the third node, which lists various notification preferences, are ranked higher than the relevant node.

ContextualRecall: The score is 0.50 because the retrieval context partially supports the expected output. The 2nd node in the retrieval context confirms that users do not receive notifications about their own activity by default, but there is no mention of the option to disable this filter in the 'Options de filtrage avancées' section of the user profile notification settings.

CustomContextualRelevancy: The retrieval context mentions filtering own events and advanced filtering options but lacks details on default notification settings and explicit user targeting.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7205128205128205

Individual_scores:

AnswerRelevancy: 0.7692307692307693

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.5

Reasons:

AnswerRelevancy: The score is 0.77 because while the answer provides some relevant information, it includes multiple irrelevant URLs that do not directly address the specific steps needed to disable notifications.

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 0.33 because the relevant node (third node) is ranked lower than irrelevant nodes. For example, the first node discusses 'how to get a notification RSS feed and the notifications macro,' and the second node lists 'various notification preferences macros but does not provide specific instructions on how to disable notifications,' both of which should be ranked lower.

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

CustomContextualRelevancy: The retrieval context mentions auto watch preferences and custom filters preferences, but lacks details on 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: 0.38181818181818183

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 answer is mostly relevant and addresses the configuration of OpenID Connect for automatic user account activation. However, it includes an irrelevant statement about the XWiki RESTful API, which slightly detracts from the overall relevance.

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

ContextualPrecision: The score is 0.00 because all the nodes in the retrieval context are irrelevant. For example, the first node provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed to automatically activate user accounts. Similarly, the second node discusses container authentication but lacks the specific configuration details needed for automatic user activation.

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 specific instructions about modifying the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.6366666666666667

Individual_scores:

AnswerRelevancy: 0.85

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.85 because the answer provides relevant information about customizing user profile fields from an OpenID Connect provider in XWiki, but includes several irrelevant links to different XWiki documentation pages 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 0.33 because the first two nodes discuss 'various authentication methods' and 'general information about the OpenID Connect authenticator', which are not relevant to mapping user profile fields. The third node, which is relevant, is ranked lower. Additionally, the fourth and fifth nodes discuss 'various types of references' and 'naming strategies for pages', which are also irrelevant and should be ranked lower.

ContextualRecall: The score is 1.00 because the retrieval context perfectly aligns with the expected output, covering all key points without any discrepancies. Great job!

CustomContextualRelevancy: The retrieval context does not contain any information about customizing user profile fields or mapping them in the `xwiki.properties` file.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.6966666666666667

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 0.9

ContextualPrecision: 0.5

ContextualRecall: 0.4666666666666667

CustomContextualRelevancy: 0.7

Reasons:

AnswerRelevancy: The score is 0.92 because the answer is mostly relevant and helpful, but it includes a link to the XWiki documentation on creating new users, which is not directly related to configuring group synchronization with an OpenID Connect provider.

Faithfulness: The score is 0.90 because the actual output incorrectly suggests setting up manual synchronization, whereas the retrieval context clearly states that the OpenID Connect Authenticator performs automatic synchronization.

ContextualPrecision: The score is 0.50 because the second node, which is highly relevant as it discusses the default group synchronization settings and the `oidc.groups.mapping` property, is ranked second. However, irrelevant nodes such as the first node, which only provides general information about prerequisites and installation instructions, and the third node, which provides general information about the OpenID Connect Authenticator, are ranked higher than they should be.

ContextualRecall: The score is 0.47 because while the 2nd node in the retrieval context mentions key properties like `xwiki.properties`, `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, it lacks specific instructions and examples provided in the expected output, such as opening the file with a text editor, setting properties with actual values, and saving and restarting the instance.

CustomContextualRelevancy: The retrieval context mentions configuring group synchronization in the xwiki.properties file and setting properties like oidc.groups.mapping, but lacks detailed steps and examples found in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_result.json

Overall_score: 0.3529411764705882

Individual_scores:

AnswerRelevancy: 0.7647058823529411

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.76 because the answer provides some relevant information but includes multiple irrelevant statements and links that do not directly address the specific question about limiting file types and sizes.

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 configuring file upload restrictions. For example, the first node discusses 'macros and preferences related to notifications, office document viewing, and PDF table of contents,' which are not pertinent to the input query.

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

CustomContextualRelevancy: The retrieval context does not contain any information about 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: 0.7963636363636363

Individual_scores:

AnswerRelevancy: 0.18181818181818182

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.18 because the majority of the statements provided are irrelevant to the question about managing script rights for a space administrator in XWiki, focusing instead on unrelated topics like Firefox script settings. However, the score is not zero because there may be a minimal connection or attempt to address script management, albeit incorrectly.

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 are ranked higher than irrelevant nodes, ensuring the most pertinent information is prioritized. Great job!

ContextualRecall: The score is 1.00 because the expected output is fully supported by the nodes in the retrieval context, indicating perfect alignment.

CustomContextualRelevancy: The retrieval context mentions that rights implied by an admin right cannot be overridden by a denial at a lower level, aligning with the expected output that admin rights cannot be refused. However, it does not explicitly state that this is not supported in XWiki.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_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 main response is relevant and helpful, but the inclusion of a URL that does not directly address the configuration question about team spaces and permissions in XWiki slightly detracts from the overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect faithfulness!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node explains the technical details of how spaces and pages are created and managed in XWiki, but it doesn't provide any information on how to set access controls for different teams. Similarly, the second node lists various macros and their functionalities in XWiki, which is unrelated to configuring access controls 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, defining rights at the wiki or space level, or inheriting rights for child pages.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.5155555555555555

Individual_scores:

AnswerRelevancy: 0.7777777777777778

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.78 because the answer provides some relevant information about authentication, but includes multiple URLs that do not directly address the specific question about authenticating users with access tokens.

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

ContextualPrecision: The score is 0.50 because the first node discusses claims accessed via the UserInfo endpoint, which is not about authentication with access tokens, and should be ranked lower. The second node is correctly ranked as it provides relevant information about configuring authorized applications. The third node provides general information about permissions and extension ratings, which is not relevant and should be ranked lower. The fourth node discusses security settings and authentication strategies for failed login attempts, also not relevant and should be ranked lower. The fifth node discusses OpenID Connect claims and group synchronization, which is not directly related and should be ranked lower.

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 mentions the 'Authorized Applications' configuration and some aspects of token claims, but it lacks detailed information about JWT, Ed25519 key, and specific steps for enabling and configuring the token authenticator.
Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_032_result.json

Overall_score: 0.4
Individual_scores:
AnswerRelevancy: 1.0
Faithfulness: 1
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:
AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and there are no irrelevant statements.
Faithfulness: The score is 1.00 because there are no contradictions. Great job on maintaining perfect alignment with the retrieval context!
ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the question about authenticating users with access tokens. For example, the first node discusses 'group synchronization and customization for OpenID Connect,' and the second node is about 'protection levels for cookies and security strategies,' neither of which are related to the input question.
ContextualRecall: The score is 0.00 because none of the sentences in the expected output can be attributed to any nodes in the retrieval context.
CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication for LLM applications, JWT tokens, or the specific configuration steps mentioned in the expected output.
Question_language: de
Expected_answer_language: de
Answer_language: en

File: qa_033_result.json

Overall_score: 0.3529411764705882
Individual_scores:
AnswerRelevancy: 0.7647058823529411
Faithfulness: 1.0
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.0
Reasons:
AnswerRelevancy: The score is 0.76 because while the response partially addresses the question about authenticating users with access tokens, it includes several URLs that are not directly relevant to the specific topic of using access tokens for authentication.
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. For example, the first node 'describes properties and methods related to collections and their management,' which is not related 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 about authenticating users with access tokens, 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_mini-128k_4b_Q4

File: qa_001_result.json

Overall_score: 0.8

Individual_scores:

AnswerRelevancy: 0.6666666666666666

Faithfulness: 1

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.67 because the answer includes irrelevant information about Alice and Ashton Smith, which does not address the issue of fixing the 'Bell' notification in XWiki. However, the relevant parts of the answer do provide some useful context related to the 'Alert Notification' extension.

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 third rank mentions the `notifications.enabled` setting in `xwiki.properties`, which directly addresses the solution. However, irrelevant nodes in the first, second, fourth, and fifth ranks discuss topics like notification RSS feed, overriding templates, 'Clear All' link, and bell icon display, which do not address enabling notifications in the `xwiki.properties` file.

ContextualRecall: The score is 1.00 because the retrieval context perfectly supports the expected output. Great job!

CustomContextualRelevancy: The retrieval context includes information about enabling notifications by setting `notifications.enabled` to `true` in `xwiki.properties`, which matches the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.36119373776908026

Individual_scores:

AnswerRelevancy: 0.0273972602739726

Faithfulness: 1

ContextualPrecision: 0.25

ContextualRecall: 0.42857142857142855

CustomContextualRelevancy: 0.1

Reasons:

AnswerRelevancy: The score is 0.03 because the output contains numerous nonsensical and irrelevant statements that do not address the issue of not receiving notifications in XWiki.

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, which mentions the 'notifications.enabled' setting in 'xwiki.properties', is ranked fourth. Irrelevant nodes, such as the first node discussing the 'Clear All' link and default behavior for own events, the second node listing dependencies for the notifications extension, and the third node providing various ways to receive notifications without addressing enabling notifications, are ranked higher.

ContextualRecall: The score is 0.43 because while the retrieval context covers enabling notifications in the 'xwiki.properties' file and mentions the alert entry represented by a bell, it lacks information on subscribing to pages or users, following users, and the network tab.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties but lacks details on subscribing to pages or users, and how to manage these subscriptions.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.48

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.50 because the output contains incomplete and irrelevant statements that do not address the query about enabling notifications for one's own actions in XWiki. However, it is not lower because there may still be some relevant information present.

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. For example, the first node discusses how to get a notification RSS feed and general notification settings, but it does not address enabling notifications for one's own actions.

ContextualRecall: The score is 0.50 because the retrieval context mentions the ability to enable or disable notification types (node 1), which relates to advanced filtering options in user settings (sentence 2). However, it does not cover that users do not receive notifications about their own activity unless explicitly targeted (sentence 1).

CustomContextualRelevancy: The retrieval context mentions notification settings and default behaviors but lacks specific details about disabling 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: 0.5066666666666666

Individual_scores:

AnswerRelevancy: 0.3333333333333333

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.33 because the actual output contains multiple irrelevant statements that do not address the question about disabling notifications in XWiki. However, it is not lower because there might still be some relevant information present in the output.

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. For example, the first node discusses enabling and disabling the email feature and live email notifications grace time, but it does not address how to disable notifications for specific pages or user settings. Similarly, the second node talks about RSS feeds for notifications and the Notifications Macro but does not provide information on disabling notifications for specific pages or user settings.

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

CustomContextualRelevancy: The retrieval context mentions settings and filters but lacks details about automatic page watching and deleting filters.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_result.json

Overall_score: 0.3142857142857143

Individual_scores:

AnswerRelevancy: 0.5714285714285714

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.57 because the output contains several irrelevant statements that do not address the specific configuration question about OpenID Connect in XWiki. However, the score is not lower because some parts of the output may still contain useful information related to the topic.

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 general features and installation details of the OpenID Connect Authenticator, which is not specific to enabling user accounts upon first login. The second node is about container authentication and only mentions OpenID Connect in passing. The third node discusses various authentication methods for the XWiki RESTful API without providing specific configuration instructions. The fourth node is about setting access rights for terminal pages, and the fifth node is about logging in and retrieving forgotten usernames or passwords, both of which are unrelated to the input query.

ContextualRecall: The score is 0.00 because the retrieval context does not mention adjusting the `xwiki.properties` file, enabling user accounts upon first login using OpenID Connect, or provide any related instructions.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.22105263157894733

Individual_scores:

AnswerRelevancy: 0.10526315789473684

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.11 because the majority of the statements in the actual output are irrelevant or nonsensical, and do not address the question about customizing user profile fields from an OpenID Connect provider in XWiki.

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. For instance, the first node discusses the OpenID Connect authenticator, its developer, active installs, rating, license, and compatibility but does not mention customizing user profile fields or mapping additional fields. Similarly, the second node explains various authentication methods for the XWiki RESTful API but does not provide information on customizing user profile fields or mapping additional fields from the OpenID Connect provider. These irrelevant nodes should be ranked lower.

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 customizing user profile fields or mapping them in the xwiki.properties file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.3807692307692308

Individual_scores:

AnswerRelevancy: 0.2

Faithfulness: 1

ContextualPrecision: 0.25

ContextualRecall: 0.15384615384615385

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.20 because the output contains multiple irrelevant and incoherent statements that do not address the configuration of 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 0.25 because the relevant node discussing 'group synchronization, including the use of the `oidc.groups.mapping` property and the `xwiki.properties` file' is ranked fourth. Irrelevant nodes, such as the first node providing 'general information about the OpenID Connect Authenticator' and the second node with 'links to various configuration guides and examples for different systems,' should be ranked lower.

ContextualRecall: The score is 0.15 because the retrieval context only partially matches the expected output, specifically in configuring group synchronization and defining group mappings, but lacks detailed steps and examples provided in the expected output.
CustomContextualRelevancy: The retrieval context mentions group synchronization and properties like `oidc.groups.mapping`, but lacks detailed steps and examples found in the expected output.
Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_008_result.json

Overall_score: 0.7909090909090908
Individual_scores:
AnswerRelevancy: 0.8
Faithfulness: 1.0
ContextualPrecision: 1.0
ContextualRecall: 0.454545454545453
CustomContextualRelevancy: 0.7
Reasons:

AnswerRelevancy: The score is 0.80 because the answer partially addresses the question but includes irrelevant information about XWiki syntax documentation, WordPress, and general statements that do not provide specific steps for setting upload restrictions.

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

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

ContextualRecall: The score is 0.45 because while the retrieval context covers the steps for configuring the maximum attachment size (Steps 2.1 to 2.5), it does not provide any information on configuring allowed mimetypes (Step 1).

CustomContextualRelevancy: The retrieval context includes steps for setting maximum attachment size and mentions mimetype restrictions, but lacks specific details on configuring the ``xwiki.properties`` file and the exact mimetype `allowList`.

Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_009_result.json

Overall_score: 0.32
Individual_scores:
AnswerRelevancy: 0.5
Faithfulness: 1
ContextualPrecision: 0
ContextualRecall: 0.0
CustomContextualRelevancy: 0.1
Reasons:

AnswerRelevancy: The score is 0.50 because while some parts of the output might be related to the topic of script rights, there are significant irrelevant statements that 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 0.00 because all nodes in the retrieval context are irrelevant to the input. For instance, the first node discusses methods for checking access rights and handling access

denial, which does not address the specific question. Similarly, the fourth node provides details on HTTP methods for managing attachments and objects in a wiki, which is unrelated to the question.
ContextualRecall: The score is 0.00 because the retrieval context does not mention or explicitly state that rights implied by admin rights cannot be denied, which is the main point of the expected output.
CustomContextualRelevancy: The retrieval context does not contain information about denying rights implied by admin right.
Question_language: en
Expected_answer_language: en
Answer_language: en

File: qa_010_result.json

Overall_score: 0.420000000000000004

Individual_scores:

AnswerRelevancy: 0.9

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.90 because the answer mostly addresses the question about configuring permissions and roles in XWiki, but includes URLs that are not directly relevant to the specific task of configuring permissions and roles.

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

ContextualPrecision: The score is 0.20 because the relevant node discussing 'setting permissions explicitly for groups or users' is ranked fifth, while the top four nodes are irrelevant, discussing topics like 'how spaces and sub-spaces are created', 'parent/child relationship and nested pages', 'various macros and their functions', and 'XWiki Syntax 2.1 Link Specification'.

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 wiki-wide rights, space-level rights, or inheriting rights for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_result.json

Overall_score: 0.23333333333333334

Individual_scores:

AnswerRelevancy: 0.16666666666666666

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.17 because the output contains multiple irrelevant and incoherent statements that do not address the issue of the missing bell icon or the Alert Notification extension.

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. For example, the first node talks about getting a notification RSS feed and settings for notifications, but does not mention enabling notifications in the wiki settings. The second node discusses overriding default notification templates and installation instructions, which are not relevant to enabling notifications in the wiki settings.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications through `xwiki.properties` or setting `notifications.enabled` to `true`, and it lacks any information about the `xwiki.properties` file or the `notifications.enabled` setting.

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

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_012_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 URLs and statements that do not address the issue of not receiving notifications.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context discuss topics unrelated to enabling notifications, such as 'various applications and extensions in XWiki' (first node), 'access rights and execution contexts' (second node), 'scripting macros and link renaming' (third node), 'document chunking and embedding models for similarity search' (fourth node), and 'configuring permissions for guest users in an open wiki' (fifth node).

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.

CustomContextualRelevancy: The retrieval context does not contain any information related to enabling notifications, following pages or users, or configuring notification settings as described in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_013_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 provided URLs and statements are entirely irrelevant and nonsensical in the context of enabling notifications for one's own activity.

Faithfulness: The score is 1.00 because there are no contradictions. 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 'JWT token generation' and the second node is about 'database migration and invalid extensions', neither of which relate to notifications or user activity settings.

ContextualRecall: The score is 0.00 because the expected output sentences about notification settings and advanced filter options do not correspond to any information in the nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context does not contain any information related to 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: 0.38076923076923075

Individual_scores:

AnswerRelevancy: 0.9038461538461539

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.90 because the answer is mostly relevant and helpful, but it includes several irrelevant URLs that do not address the question about disabling notifications.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'XWiki syntax and formatting,' which is unrelated to managing notifications. The second node describes 'an extension for user authentication using JSON Web Tokens,' which is not relevant to the question about disabling notifications. Additionally, the third node provides 'links and descriptions related to the LLM Application and its API,' which do not pertain 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 tracking of pages 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: 0.49000000000000005

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.2

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and directly addresses the question 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 0.25 because the relevant node in the retrieval context is ranked fourth, while the first three nodes are irrelevant. The first node provides general information about the OpenID Connect Authenticator, the second node offers links to configuration guides without specific instructions, and the third node discusses other authentication methods. The relevant information about modifying the 'xwiki.properties' file is not prioritized.

ContextualRecall: The score is 0.20 because only the 2nd sentence in the expected output can be related to the 2nd node in the retrieval context, while the rest of the sentences do not find any corresponding information in the retrieval context.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.68

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.2

ContextualRecall: 1.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is perfectly relevant and there are no irrelevant statements in the actual output. 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 discusses configuring user profile fields in the 'xwiki.properties' file, is ranked fifth. The irrelevant nodes, such as the first node that provides general information about prerequisites and installation instructions, and the second node that describes general limitations, should be ranked lower.

ContextualRecall: The score is 1.00 because all sentences in the expected output align perfectly with the nodes in the retrieval context, demonstrating a comprehensive match.

CustomContextualRelevancy: The retrieval context mentions 'xwiki.properties' and customization but lacks specific details on 'oidc.user.mapping' and the example provided in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.5276717557251909

Individual_scores:

AnswerRelevancy: 0.7633587786259542

Faithfulness: 0.875

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.76 because while the response contains some relevant information, it is heavily diluted with numerous irrelevant statements that do not pertain to setting up group synchronization between XWiki and an OpenID Connect provider.

Faithfulness: The score is 0.88 because the actual output incorrectly suggests using LDAP for handling user information and groups, while the retrieval context specifies that LDAP is deprecated for XWiki versions 7.4 and below, and OpenID Connect Authenticator is used for XWiki 14.10+.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node provides general information about the OpenID Connect Authenticator, but does not specifically address the setup process for group synchronization. The second node describes the OpenID Connect Authenticator's details, but lacks configuration information. The third node discusses various authentication methods without mentioning group synchronization setup. The fourth node mentions group synchronization but lacks detailed steps. The fifth node provides a table of contents without specific setup steps.

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

CustomContextualRelevancy: The retrieval context does not contain specific instructions for configuring `xwiki.properties` for group synchronization with OpenID Connect.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_018_result.json

Overall_score: 0.6690909090909092

Individual_scores:

AnswerRelevancy: 0.2

Faithfulness: 1

ContextualPrecision: 1.0

ContextualRecall: 0.5454545454545454

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.20 because the majority of the output contains irrelevant information that does not address the question about restricting file uploads to images or PDFs with a maximum size of 10 MB in a wiki. The provided links and nonsensical statements detract significantly from the relevance.

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

ContextualPrecision: The score is 1.00 because the relevant nodes are ranked higher than the irrelevant nodes. The first node provides detailed steps on setting limits on the maximum size of an attachment, which is directly relevant to configuring the maximum attachment size to 10 MB. The irrelevant nodes, such as the second node mentioning MIME type configurations without specific steps, and the third node about managing documents, are ranked lower.

ContextualRecall: The score is 0.55 because while the retrieval context provides detailed steps for configuring the maximum attachment size (sentences 6-11), it does not mention configuring MIME types or ensuring only images or PDFs can be uploaded (sentences 1-5).

CustomContextualRelevancy: The retrieval context includes steps for setting the maximum attachment size and mentions MIME type restrictions, but it does not provide the specific MIME type configuration line or the exact 10 MB size in bytes.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_019_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 that do not address the query about denying script rights to a space administrator.

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 question. For example, the first node contains links to images and references, which are not relevant to the question about denying script rights to a domain administrator.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about 'Rechte' or 'Admin-Recht' being denied or supported in XWiki, which is the main focus of the expected output.

CustomContextualRelevancy: The retrieval context does not contain any information related to rights or admin rights in XWiki.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_020_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 while the answer partially addresses the question about configuring XWiki, it contains nonsensical and irrelevant statements that detract from its overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'additional actions that can be performed in XWiki, such as creating children pages and changing page access rights,' but it does not provide specific steps for configuring access control as required by the input. Similarly, the second node explains 'the concept of wikis and panels in XWiki,' which does not address the specific access control configuration needed for the input question.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about creating groups, setting wiki-wide rights, setting area-specific rights, or inheriting rights for subpages.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_021_result.json

Overall_score: 0.20392156862745098

Individual_scores:

AnswerRelevancy: 0.0196078431372549

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.02 because the majority of the statements in the actual output are irrelevant and do not address the issue of the notification bell not being available, as highlighted by the numerous unrelated and nonsensical statements.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant, discussing topics like 'getting an RSS feed of notifications' (first node), 'overriding default notification templates' (second node), 'Clear All link and default notification behaviors' (third node), 'custom notification filters and the watch feature' (fourth node), and 'various settings related to notifications' (fifth node), none of which 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 of the nodes in the retrieval context.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_022_result.json

Overall_score: 0.4666666666666667

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.8

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.00 because the output contains numerous irrelevant and unclear statements that do not address the issue of not receiving notifications.

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

ContextualPrecision: The score is 0.33 because the relevant node in the retrieval context is ranked third. The first node discusses 'getting a notification RSS feed and the settings for notifications', which is not relevant to enabling notifications or following pages/users. The second node is about 'overriding default notification templates and installation instructions', which is also not relevant. The relevant node,

which mentions 'Follow a user' and 'Watched Entities', should be ranked higher.

ContextualRecall: The score is 0.80 because while the retrieval context covers following pages, users, and viewing notifications, it lacks information about activating notifications through the `xwiki.properties` file.

CustomContextualRelevancy: The retrieval context mentions following users and enabling notifications 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: en

File: qa_023_result.json

Overall_score: 0.3672727272727273

Individual_scores:

AnswerRelevancy: 0.03636363636363636

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.04 because the output contains numerous irrelevant and convoluted statements 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 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. For example, the first node discusses how to get a notification RSS feed and does not address notifications for one's own activity, while the second node provides a list of various notification-related links and settings but does not specifically address notifications for one's own activity.

ContextualRecall: The score is 0.50 because the 1st sentence in the expected output can be linked to the 1st node in the retrieval context, but the 2nd sentence cannot be attributed to any node in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions filtering own events and notification settings, but lacks details on disabling the filter in 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_024_result.json

Overall_score: 0.6433333333333333

Individual_scores:

AnswerRelevancy: 0.3333333333333333

Faithfulness: 1

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.75

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.33 because the output contains multiple irrelevant statements that do not address the question about disabling notifications, but it does contain some relevant information.

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 (rank 3) discusses 'Auto Watch' and 'Default Watch Settings,' which are directly related to managing automatic notifications. However, irrelevant nodes ranked higher (ranks 1, 2, 4, and 5) discuss topics such as notification RSS feeds, notification preferences macros, and notification templates, which are not related to disabling notifications or modifying settings.

ContextualRecall: The score is 0.75 because the retrieval context nodes support the first three sentences of the expected output, but there is no information about 'Filtres personnalisés' in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions auto-watched pages and notification settings but lacks explicit details about modifying the 'Suivi automatique des pages' and deleting custom filters.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_025_result.json

Overall_score: 0.38

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.4

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.50 because while the answer addresses the configuration of OpenID Connect, it includes irrelevant information about the XWiki RESTful API documentation and unnecessary instructions, which do not directly help with the specific issue of automatically activating user accounts.

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 describes the OpenID Connect Authenticator extension but lacks information on configuring user account activation. The second node discusses container authentication and the xwiki.cfg file without addressing OpenID Connect or user account activation. The third node provides a general description of the OpenID Connect Authenticator and links to various configuration guides but does not mention the specific configuration needed for automatic user activation. The fourth node explains different authentication methods for the XWiki RESTful API and does not address the configuration of OpenID Connect or user account activation. The fifth node lists various endpoints and customization options for the OpenID Connect Authenticator but does not provide information about configuring automatic user account activation.

ContextualRecall: The score is 0.40 because while the retrieval context nodes mention configuring OpenID Connect in XWiki and imply automatic activation of user accounts, they do not provide specific instructions for editing the `xwiki.properties` file, setting the `oidc.enableUser` property, or restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the xwiki.properties file or setting the oidc.enableUser property.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_026_result.json

Overall_score: 0.43916666666666665

Individual_scores:

AnswerRelevancy: 0.0625

Faithfulness: 1

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.8

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.06 because the majority of the statements in the actual output are irrelevant to the question about customizing user profile fields from an OpenID Connect provider in XWiki.

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, which discusses configuring OpenID Connect endpoints in the `xwiki.properties` file, is ranked third. Irrelevant nodes discussing various authentication methods (first node) and general information about the OpenID Connect authenticator (second node) should be ranked lower.

ContextualRecall: The score is 0.80 because the retrieval context aligns well with the sentences about mapping user profile fields and provides relevant examples, but it does not mention saving the `xwiki.properties` file or restarting XWiki.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_027_result.json

Overall_score: 0.6360683760683761

Individual_scores:

AnswerRelevancy: 0.1111111111111111

Faithfulness: 1

ContextualPrecision: 0.5

ContextualRecall: 0.7692307692307693

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.11 because the majority of the statements in the actual output are irrelevant to the input question about configuring group synchronization between XWiki and an OpenID Connect provider. Only a very small portion of the content might be tangentially related, hence the low but non-zero 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 0.50 because the relevant node (rank 2) that mentions 'default group synchronization' and 'properties in the `xwiki.properties` file' is ranked lower than several irrelevant nodes. For example, the first node provides 'general information about prerequisites, installation, and dependencies' and should be ranked lower.

ContextualRecall: The score is 0.77 because while the retrieval context (especially the 2nd node) provides detailed information on configuring properties such as `oidc.groups.claim`, `oidc.groups.separator`, and `oidc.groups.mapping`, it does not explicitly mention steps like locating, opening, and saving the `xwiki.properties` file or restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context contains information about configuring group synchronization in the xwiki.properties file, including the use of oidc.groups.mapping and

oidc.groups.claim properties. However, it lacks specific details about setting oidc.groups.separator and saving and restarting the instance.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_028_result.json

Overall_score: 0.2

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.00 because the output contains numerous irrelevant and incoherent statements that do not address the question about ensuring only image or PDF files of a maximum size of 10 MB can be uploaded.

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 discuss topics such as 'macros and functionalities related to notifications, office document viewing, and PDF table of contents' (first node), 'customizing the PDF cover by overriding a template and adding a company logo' (second node), 'creating a new page by adding a wanted link and using templates' (third node), 'macros related to navigation, content display, and notifications' (fourth node), and 'CSS style information for images' (fifth node), which are irrelevant to configuring file upload restrictions.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about configuring MIME types or setting maximum upload size for attachments.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_029_result.json

Overall_score: 0.7

Individual_scores:

AnswerRelevancy: 0.5

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 1.0

Reasons:

AnswerRelevancy: The score is 0.50 because the provided link and statement do not address how to deny script rights to a space administrator, making the information partially irrelevant.

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. For example, the first node discusses 'default right settler additional policies,' which does not address the question about denying script rights to an administrator. The second node talks about 'executing tasks

with user rights,' which is also not relevant. Similarly, the third node mentions the 'Scripting API and checking access rights,' but does not address the specific question. These irrelevant nodes should be ranked lower.

ContextualRecall: The score is 1.00 because the expected output is fully supported by the retrieval context, perfectly matching the information provided.

CustomContextualRelevancy: The retrieval context clearly states that rights implied by an admin right cannot be overridden by a denial at a lower level, matching the expected answer that admin rights cannot be refused.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

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 the output contains multiple irrelevant statements, including URLs to unrelated pages, incomplete and nonsensical sentences, and instructions that do not address the question about configuring team spaces in XWiki.

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 configuring access controls for teams. For example, the first node describes how user-created pages and spaces work in XWiki, but it does not provide any information on how to set access controls for teams. Similarly, the second node provides details on various macros and scripting languages supported by XWiki, which is not relevant to configuring access controls for teams.

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.

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.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

File: qa_031_result.json

Overall_score: 0.36857142857142855

Individual_scores:

AnswerRelevancy: 0.0

Faithfulness: 1

ContextualPrecision: 0.5

ContextualRecall: 0.14285714285714285

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.00 because the output contains multiple irrelevant URLs and statements that do not address the question about authenticating users with access tokens.

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

ContextualPrecision: The score is 0.50 because the first node discusses claims at the UserInfo endpoint, which is not relevant to authenticating users with access tokens, and should be ranked lower. The second node is relevant as it provides information about configuring 'Authorized Applications'. The third node, providing general information about the XWiki Security Module, is also irrelevant and should be ranked lower. The fourth node discusses cookie protection and security strategies, which are not relevant, and the fifth node discusses custom claims and group synchronization, which is also not directly relevant.

ContextualRecall: The score is 0.14 because only the 2nd sentence in the expected output can be attributed to the 2nd node in the retrieval context, while the rest of the sentences lack corresponding information in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions authorized applications and token claims but lacks details on enabling the token authenticator, JWT token generation, and specific claims required.

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

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.50 because the actual output contains irrelevant statements that do not address the input question about authenticating users with access tokens, but it may still contain some relevant information.

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 discusses 'group synchronization and customization for OpenID Connect,' which is not related to authenticating users with access tokens. Similarly, the second node is about 'protection levels for cookies and security strategies,' which does not pertain to the use of access tokens for user 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 for the LLM application (BETA) extension or the specific steps and requirements for using JWT tokens.

Question_language: de

Expected_answer_language: de

Answer_language: en

File: qa_033_result.json

Overall_score: 0.21538461538461537

Individual_scores:

AnswerRelevancy: 0.07692307692307693

Faithfulness: 1

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.08 because the majority of the statements in the actual output are irrelevant to the question about authenticating users with access tokens, including unrelated URLs and nonsensical phrases.

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 'properties and methods related to a collection resource,' which is not relevant to user authentication with access tokens. Similarly, the second node discusses 'various claims at the UserInfo endpoint,' which does not provide information on authenticating users with access tokens.

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

CustomContextualRelevancy: The retrieval context does not contain any information about authenticating users with access tokens or JWT, nor details about configuring the authenticator in xwiki.properties.

Question_language: fr

Expected_answer_language: fr

Answer_language: en

Model: *AI.Models.qa_qwen2_7b_Q4*

File: *qa_001_result.json*

Overall_score: 0.7895238095238095

Individual_scores:

AnswerRelevancy: 0.7142857142857143

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.9

Reasons:

AnswerRelevancy: The score is 0.71 because the headings provided in the output do not offer actionable information directly, but the rest of the content is relevant and addresses 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 0.33 because the relevant node (rank 3) that mentions the 'notifications.enabled' setting in 'xwiki.properties' is ranked lower than several irrelevant nodes. For example, the first node discusses how to get a notification RSS feed and the settings for notifications, but it does not address enabling notifications in 'xwiki.properties'. Additionally, the second node talks about overriding default notification templates and installation instructions, which are not relevant to enabling notifications in 'xwiki.properties'.

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

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties but does not explicitly state setting it to 'true'.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_002_result.json

Overall_score: 0.54

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.25

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 1.00 because the answer is fully relevant and there are no irrelevant statements in the actual output. 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 0.25 because the relevant node (rank 4) is ranked lower than several irrelevant nodes. For example, the first node discusses the 'Clear All' link, grouping notifications, and notification strategies, but does not address enabling notifications or subscribing to pages/users. Additionally, the second node lists dependencies for the notifications extension, which is not relevant to the question of why notifications are not being received.

ContextualRecall: The score is 0.25 because only the instruction to enable notifications by setting `notifications.enabled` to `true` in `xwiki.properties` (sentence 2) is supported by the 5th node in the retrieval context, while other instructions related to subscribing to pages or users, watching locations, and following users are not supported by any nodes in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions enabling notifications in xwiki.properties and the alert menu but lacks details on subscribing to pages or users and viewing the network tab in the user profile.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_003_result.json

Overall_score: 0.45999999999999996

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.3

Reasons:

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

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. The first node is about getting an RSS feed of notifications, which does not address how to enable notifications for one's own actions. The second node lists various notification settings options but does not mention how to enable notifications for one's own actions. The third node discusses enabling and disabling the email feature at the platform level, which is not relevant. The fourth node describes different ways to

receive notifications in XWiki but does not address enabling notifications for one's own actions. The fifth node is about overriding default notification templates and installation instructions, which is also not relevant.

ContextualRecall: The score is 0.00 because the retrieval context does not mention that users don't receive notifications about their own activity unless explicitly targeted, nor does it mention the ability to disable this filter in the 'Advanced filtering options' section of the user's notification settings.

CustomContextualRelevancy: The retrieval context mentions default notification settings and the ability to enable/disable notifications, but lacks specific details about disabling the filter for one's own activity in the 'Advanced filtering options' section.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_004_result.json

Overall_score: 0.6133333333333334

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0.3666666666666667

ContextualRecall: 0.5

CustomContextualRelevancy: 0.2

Reasons:

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

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

ContextualPrecision: The score is 0.37 because the relevant nodes in the retrieval context are ranked lower than irrelevant nodes. For example, the first node discusses enabling and disabling the notification module and email notifications at the platform level, which is not directly related to user-specific notification settings. Similarly, the second node talks about getting an RSS feed of notifications and the general settings menu but does not provide specific information about disabling notifications for pages the user is not interested in. These irrelevant nodes are ranked higher than relevant nodes, such as the third node which mentions 'Watch and custom filters' and 'Enabling notifications for all users by default,' and the fifth node which includes 'Choosing the default auto-watched mode,' both directly addressing the user's issue.

ContextualRecall: The score is 0.50 because while some aspects of notifications and settings were mentioned in the nodes in retrieval context, key details about being notified about changes on your work and changing the 'Automatic page watching' setting were not explicitly covered.

CustomContextualRelevancy: The retrieval context mentions settings and filters but lacks details on automatic page watching and deleting filters created this way.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_005_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 answer is highly relevant and addresses the question directly, but it includes a link to the XWiki RESTful API documentation, which is not directly related to configuring OpenID Connect authentication.

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 provides general information about the OpenID Connect Authenticator but does not address the specific configuration needed to enable user accounts upon first login. The second node discusses container authentication and mentions OpenID Connect in passing, but it does not provide the specific configuration details required to solve the problem. The third node explains various authentication methods, including OpenID Connect, but does not provide the specific configuration needed to automatically enable user accounts upon first login. The fourth node is about setting access rights for a terminal page and does not relate to configuring OpenID Connect authentication. The fifth node is about logging in and retrieving forgotten usernames and passwords, which is not relevant to the configuration of OpenID Connect to enable user accounts upon first login.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about adjusting the `xwiki.properties` file or the `oidc.enableUser` property.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_006_result.json

Overall_score: 0.3408199643493761

Individual_scores:

AnswerRelevancy: 0.9393939393939394

Faithfulness: 0.7647058823529411

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.94 because the answer is mostly relevant and addresses the question about customizing user profile fields from an OpenID Connect provider in XWiki. However, it includes links to the XWiki RESTful API documentation and the Administration Application extension, which are not directly relevant to the specific query.

Faithfulness: The score is 0.76 because the actual output contains several contradictions: it mentions `config.properties` or `config.ini` for setting up a connection, `xwiki.properties` or the XWiki administration interface for defining custom user properties, and an `AuthenticationHandler` class for user authentication and authorization, none of which are mentioned in the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For instance, the first node discusses the OpenID Connect Authenticator extension but does not provide any information related to customizing user profile fields or mapping additional fields. Similarly, the second node explains various authentication methods for the XWiki RESTful API but does not address the customization of user profile fields or mapping additional fields from the 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 does not contain any information about customizing user profile fields or mapping them in the xwiki.properties file.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_007_result.json

Overall_score: 0.554317617866005

Individual_scores:

AnswerRelevancy: 0.967741935483871

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.15384615384615385

CustomContextualRelevancy: 0.4

Reasons:

AnswerRelevancy: The score is 0.97 because the answer is highly relevant and provides a detailed explanation for configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes an irrelevant link to the XWiki RESTful API documentation, which slightly detracts from the overall relevance.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness to the retrieval context!

ContextualPrecision: The score is 0.25 because the relevant node that directly addresses group synchronization (rank 4) is ranked lower than several irrelevant nodes. For example, the first node provides general information about the OpenID Connect Authenticator without addressing group synchronization, and the second node describes installation and general configuration without specific steps for group synchronization.

ContextualRecall: The score is 0.15 because the retrieval context only partially supports the expected output. Specifically, only sentence 3 and sentence 7 in the expected output are somewhat supported by node 6 in the retrieval context, while the rest of the sentences lack corresponding information in the retrieval context.

CustomContextualRelevancy: The retrieval context mentions group synchronization and some properties in xwiki.properties but lacks detailed steps and examples found in the expected output.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_008_result.json

Overall_score: 0.8218181818181819

Individual_scores:

AnswerRelevancy: 0.9090909090909091

Faithfulness: 1.0

ContextualPrecision: 1.0

ContextualRecall: 0.5

CustomContextualRelevancy: 0.7

Reasons:

AnswerRelevancy: The score is 0.91 because the majority of the response is relevant and helpful, but it includes repeated URLs that do not provide the necessary information on how to ensure only images or PDF files of a maximum 10MB can be uploaded.

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, which provides detailed steps on setting the maximum size of an attachment, is ranked first. Irrelevant nodes, such as those about moving attachments (rank 2), whitelists and blacklists (rank 3), linking to external files (rank 4), and various linking methods (rank 5), are correctly ranked lower.

ContextualRecall: The score is 0.50 because while the retrieval context nodes provide detailed steps for configuring the maximum attachment size (Step 2, sentences 1-5), they do not mention the restriction to images or PDF files, configuring allowed mimetypes, or editing the `xwiki.properties` file (Step 1, sentences 1-3).

CustomContextualRelevancy: The retrieval context contains information about setting the maximum attachment size and mentions mimetype restrictions, but it lacks specific details about configuring the `xwiki.properties` file and the exact mimetype allow list.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_009_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 answer mostly addresses the question but includes irrelevant information such as a URL that does not directly address the issue and a statement about Node.js and Express that is not relevant to the platform in question.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining perfect faithfulness!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node 'describes methods for checking access rights but does not address the specific question of denying script rights to a space administrator', and the fourth node 'describes HTTP methods for managing attachments in a wiki, which is unrelated to denying script rights to a space administrator'.

ContextualRecall: The score is 0.00 because the retrieval context does not mention that certain rights cannot be denied if implied by admin rights, nor does it state that rights implied by admin rights cannot be denied.

CustomContextualRelevancy: The retrieval context does not mention that rights implied by admin right cannot be denied.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_010_result.json

Overall_score: 0.43966386554621845

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 0.9411764705882353

ContextualPrecision: 0.2

ContextualRecall: 0.0

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.86 because the answer mostly addresses the configuration steps needed in XWiki, but includes irrelevant details such as the URL and example team names that do not directly help with the configuration.

Faithfulness: The score is 0.94 because the actual output incorrectly claims that XWiki uses roles to manage permissions across spaces, which is not mentioned in the retrieval context.

ContextualPrecision: The score is 0.20 because the relevant node (rank 5) is ranked lower than four irrelevant nodes. For example, the first node 'explains how spaces and pages are created in XWiki but does not provide information on configuring access rights', and the second node 'discusses the parent/child relationship and the nested pages concept in XWiki, but it does not address access control 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 mentions setting rights and creating spaces, but lacks specific details on creating groups, setting wiki-wide rights, and inheriting rights for child pages.

Question_language: en

Expected_answer_language: en

Answer_language: en

File: qa_011_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 addresses the issue 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 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses 'getting a notification RSS feed and settings for notifications,' which does not address the availability of the notification bell. Similarly, the second node talks about 'overriding default notification templates and installation instructions,' which are not related to the notification bell issue.

ContextualRecall: The score is 0.00 because the retrieval context does not mention enabling notifications via `xwiki.properties` or any similar setting, nor does it provide information about setting `notifications.enabled` to `true` in `xwiki.properties`.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_012_result.json

Overall_score: 0.3473684210526316

Individual_scores:

AnswerRelevancy: 0.7368421052631579

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.74 because the output contains several irrelevant links that do not address the issue of not receiving notifications. However, the score is not lower because the relevant parts of the output do attempt to address the notification 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. For example, the first node discusses 'various applications like Meeting Application, Forum Application, and Ideas Application,' which are not related to notifications settings. Similarly, the second node talks about 'access rights and executing tasks with specific user rights,' which is unrelated to enabling or configuring 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 enabling notifications, following pages or users, or configuring notification settings.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_013_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 answer is mostly relevant but repeatedly includes a URL that does not address the specific question about enabling notifications for one's own 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 0.00 because all nodes in the retrieval context are unrelated to the input. For example, the first node discusses 'token generation and JWT claims', which is unrelated to notifications or user activity settings.

ContextualRecall: The score is 0.00 because none of the nodes in the retrieval context mention notifications about user activity, related settings, 'Erweiterte Filteroptionen', or user notification settings in user profiles.

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_014_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 actual output contains several irrelevant URLs that do not address the question about disabling notifications. However, the core information provided is still somewhat relevant to the input query.

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 unrelated to the input. For example, the first node discusses 'XWiki syntax and formatting features,' the second node is about 'an extension for authenticating users with JSON Web Tokens,' and the third node provides 'links and descriptions for managing collections and documents via REST API in XWiki.' These topics do not 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: 0.6209523809523809
Individual_scores:
AnswerRelevancy: 0.9047619047619048
Faithfulness: 1.0
ContextualPrecision: 0.2
ContextualRecall: 1.0
CustomContextualRelevancy: 0.0
Reasons:

AnswerRelevancy: The score is 0.90 because the answer is mostly relevant and provides useful information, but it includes some irrelevant statements that do not directly address the input question.

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

ContextualPrecision: The score is 0.20 because the relevant node in the fifth position mentions the 'xwiki.properties' file, which is pertinent to the required configuration change for automatically activating user accounts with OpenID Connect. However, the irrelevant nodes ranked higher provide general information about OpenID Connect Authenticator, general configuration guides, and various configuration options without addressing the specific property 'oidc.enableUser' that needs to be set to 'true' in 'xwiki.properties'.

ContextualRecall: The score is 1.00 because all sentences in the expected output are fully supported by the information in the retrieval context nodes, particularly the 2nd node under 'Configuration files'. Great job!

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

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_016_result.json

Overall_score: 0.56

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.6

CustomContextualRelevancy: 0.2

Reasons:

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

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining faithfulness!

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node provides general information about prerequisites, installation instructions, and dependencies for the OpenID Connect Authenticator but does not address customizing user profile fields or mapping additional fields. Similarly, the second node describes the OpenID Connect Authenticator's general functionality and provides links to various resources, but it does not mention customizing user profile fields or mapping additional fields.

ContextualRecall: The score is 0.60 because while the retrieval context nodes mention 'Customization', 'xwiki.properties', and restarting XWiki (supporting sentences 1, 2, and 5), they lack specific examples and details on mapping custom properties and replacing placeholder variables (missing support for sentences 3 and 4).

CustomContextualRelevancy: The retrieval context mentions configuration files and customization but lacks specific details on mapping user profile fields in xwiki.properties and the exact example provided in the expected output.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_017_result.json

Overall_score: 0.6501398601398602

Individual_scores:

AnswerRelevancy: 0.9545454545454546

Faithfulness: 1.0

ContextualPrecision: 0.25

ContextualRecall: 0.8461538461538461

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.95 because the overall response is highly relevant and addresses the question well, but it includes a redundant statement about 'Konfiguration des Anbieters zur XWiki' which slightly detracts from the conciseness.

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 mentioning 'Group synchronization' and linking to a section with likely relevant information is ranked fourth. The irrelevant nodes, such as the first node which provides general information about OpenID Connect and XWiki without specific steps, should be ranked lower.

ContextualRecall: The score is 0.85 because while the retrieval context covers most aspects of configuring group synchronization between XWiki and an OpenID Connect provider, it lacks explicit instructions for opening the 'xwiki.properties' file with a text editor (sentence 2) and saving the file and restarting the XWiki instance (sentence 7).

CustomContextualRelevancy: The retrieval context mentions group synchronization but lacks detailed steps and specific 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: 0.8606593406593406

Individual_scores:

AnswerRelevancy: 0.8571428571428571

Faithfulness: 0.8461538461538461

ContextualPrecision: 1.0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.86 because the answer is mostly relevant to the question about restricting file types and sizes in the wiki. However, it includes links to general documentation, an extension page, and a user guide on document lifecycle, which do not directly address the specific configuration question.

Faithfulness: The score is 0.85 because the actual output incorrectly states that only certain file types like images or PDFs with a maximum size of 10 Megabytes can be uploaded, while the retrieval context specifies that the maximum size of an attachment is set to 100GB by default and 32MB for XWiki versions less than 10.9RC1.

ContextualPrecision: The score is 1.00 because all relevant nodes are ranked higher than irrelevant nodes. For instance, the first node mentions 'setting limits on the maximum size of an attachment,' and the second node provides 'detailed steps on how to change the maximum upload size,' both of which are directly relevant. The irrelevant nodes, such as the third node discussing 'document properties and the REST API,' are correctly ranked lower.

ContextualRecall: The score is 1.00 because the retrieval context perfectly matches all the steps and details provided in the expected output. Great job!

CustomContextualRelevancy: The retrieval context includes steps to change the maximum attachment size and mentions MIME type restrictions, but lacks specific details on configuring the `xwiki.properties` file and setting the MIME type allow list.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_019_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 response addresses the question about denying script rights to a space administrator, but several irrelevant links were included, which 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. For example, the first node is a link to an image and does not provide any information related to denying script rights to a domain administrator. Similarly, the tenth node discusses Kerberos authentication and LDAP configuration, which is unrelated to the question.

ContextualRecall: The score is 0.00 because the expected output sentence does not match any part of the nodes in retrieval context, which discuss images, LDAP configurations, and form utilization.

CustomContextualRelevancy: The retrieval context does not contain any information related to the admin rights or their implications in XWiki.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_020_result.json

Overall_score: 0.3741935483870968

Individual_scores:

AnswerRelevancy: 0.8709677419354839

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.87 because the answer provides relevant information but includes multiple links that do not directly address the specific configuration 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. For example, the first node discusses 'various actions that can be performed in XWiki' but does not provide specific steps for configuring access control for teams. Similarly, the second node explains 'the concept of wikis and subwikis' without addressing the specific requirements for team access control.

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 wiki-wide rights, setting area-level rights, or inheriting rights for subordinate pages.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_021_result.json

Overall_score: 0.36923076923076925

Individual_scores:

AnswerRelevancy: 1.0

Faithfulness: 0.8461538461538461

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 1.00 because the response is perfectly relevant and there are no irrelevant statements.

Faithfulness: The score is 0.85 because the actual output incorrectly claims that you can verify if an extension is installed while offline, and it mentions a bell icon configuration that is not supported by the retrieval context.

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant. For example, the first node discusses getting a notification RSS feed and settings for notifications, but does not address enabling notifications in the wiki settings.

ContextualRecall: The score is 0.00 because the retrieval context does not mention anything about ensuring notifications are activated in the wiki or how to activate them, nor does it provide information about defining the property `notifications.enabled` in `xwiki.properties` or setting it to `true`.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_022_result.json

Overall_score: 0.6383333333333333

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.625

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is highly relevant and addresses the issue of not receiving notifications, but the inclusion of an unrelated URL slightly detracts from its 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 0.33 because the relevant node (rank 3) discussing 'Follow a user' and 'Watched Entities' is ranked lower than irrelevant nodes. For example, the first node discusses getting a notification RSS feed, settings, and enabling/disabling notification types, which is unrelated to enabling notifications or following pages/users. Similarly, the second node is about overriding default notification templates and installation instructions, which is also unrelated.

ContextualRecall: The score is 0.62 because while the retrieval context covers following pages or users and activating notifications through the alert menu, it lacks details on setting `notifications.enabled` in `xwiki.properties` and viewing the list of followed users in the 'network' tab.

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_023_result.json

Overall_score: 0.6433333333333333

Individual_scores:

AnswerRelevancy: 0.9166666666666666

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.92 because the response is mostly relevant and helpful, but it includes a link about XWiki syntax, which does not directly address the question about enabling notifications for one's own activity.

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 getting an RSS feed of notifications and does not mention anything about notifications for one's own activity. The second node lists various notification settings and filters but does not specifically address notifications for one's own activity. The third node describes different macros related to notifications but does not provide information about enabling notifications for one's own actions. The fourth node talks about overriding default notification templates and installation instructions, which are not relevant to the question about enabling notifications for one's own activity. The fifth node lists several ways to be notified in XWiki but does not specifically address how to enable notifications for one's own actions.

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

CustomContextualRelevancy: The retrieval context mentions default notification settings and filtering options but lacks details about disabling the filter in 'Options de filtrage avancées'.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_024_result.json

Overall_score: 0.7073118279569893

Individual_scores:

AnswerRelevancy: 0.9032258064516129

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 0.5

CustomContextualRelevancy: 0.8

Reasons:

AnswerRelevancy: The score is 0.90 because the main content addresses how to disable notifications, but it includes irrelevant links that do not directly pertain to the question asked.

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

ContextualPrecision: The score is 0.33 because the relevant node, which discusses 'Auto Watch' and 'Default Watch Settings' (rank 3), is ranked lower than irrelevant nodes. For example, the first node only talks about getting a notification RSS feed and settings menu, which does not help with disabling automatic notifications. Similarly, the second node lists notification preferences macros without explaining how to disable automatic notifications.

ContextualRecall: The score is 0.50 because while the retrieval context nodes mention default notification settings and accessing notification settings, they do not specifically mention being notified of changes to your work or deleting existing notification filters.

CustomContextualRelevancy: The retrieval context mentions auto watch preferences and custom notification filters, but lacks explicit details about modifying the 'Suivi automatique des pages' setting and removing existing filters.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_025_result.json

Overall_score: 0.3945945945945946

Individual_scores:

AnswerRelevancy: 0.972972972972973

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.97 because the answer is highly relevant and addresses the main question about configuring user account activation via OpenID Connect. However, it includes a minor irrelevant detail about the XWiki RESTful API, which slightly detracts from the overall relevance.

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

ContextualPrecision: The score is 0.00 because all nodes in the retrieval context are irrelevant to the input. For example, the first node 'describes the OpenID Connect Authenticator extension, its developer, and installation details, but does not mention the specific configuration needed to automatically activate user accounts.' Similarly, the second node 'discusses container authentication and mentions OpenID Connect briefly but does not provide the specific configuration details required to solve the problem.' These irrelevant nodes should be ranked lower.

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 adjusting the `xwiki.properties` file or setting the `oidc.enableUser` property.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_026_result.json

Overall_score: 0.6333333333333334

Individual_scores:

AnswerRelevancy: 0.8333333333333334

Faithfulness: 1.0

ContextualPrecision: 0.3333333333333333

ContextualRecall: 1.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.83 because while the answer provides some relevant information, it includes multiple URLs that are not directly related to customizing user profile fields from OpenID Connect providers, 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 0.33 because the relevant node (third node) is ranked lower than several irrelevant nodes. For example, the first node discusses 'various authentication methods for the XWiki RESTful API, including OIDC', which does not provide information on mapping user profile fields. Similarly, the second node provides 'general information about the OpenID Connect Authenticator for XWiki' but does not address how to map user profile fields. These irrelevant nodes should be ranked lower than the relevant node that mentions the `xwiki.properties` file and includes details about OpenID Connect endpoints, which is pertinent to configuring user profile field mappings.

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

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

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_027_result.json

Overall_score: 0.7131034482758621

Individual_scores:

AnswerRelevancy: 0.9655172413793104

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.5

CustomContextualRelevancy: 0.6

Reasons:

AnswerRelevancy: The score is 0.97 because the answer is highly relevant and mostly addresses the question about configuring group synchronization between XWiki and an OpenID Connect provider. However, it includes a link about creating new users in XWiki, which is not directly related to the group synchronization topic.

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

ContextualPrecision: The score is 0.50 because the second node, which is relevant, is ranked higher than some irrelevant nodes. However, the first, third, fourth, and fifth nodes are irrelevant as they provide general information about prerequisites, installation instructions, developer details, table of contents, and managing users and groups, but do not address the specific process of configuring group synchronization. These irrelevant nodes should be ranked lower.

ContextualRecall: The score is 0.50 because while the 2nd node in the retrieval context mentions the `xwiki.properties` file and provides examples for configuring group synchronization, it lacks specific steps such as opening the file with a text editor, replacing placeholders with actual values, and saving and restarting the XWiki instance.

CustomContextualRelevancy: The retrieval context mentions configuring group synchronization in the xwiki.properties file and using oidc.groups.mapping, but lacks detailed steps and examples provided in the expected output.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_028_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 answer is mostly relevant to the input, but it includes URLs that do not directly address the question about limiting file types and sizes for uploads.

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. For example, the first node describes 'various macros and functionalities related to notifications, office document viewing, and PDF table of contents', which are not relevant to configuring file upload restrictions.

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

CustomContextualRelevancy: The retrieval context does not contain any information about configuring MIME types or setting the maximum upload size for attachments.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_029_result.json

Overall_score: 0.6171428571428571

Individual_scores:

AnswerRelevancy: 0.7857142857142857

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 1.0

CustomContextualRelevancy: 0.3

Reasons:

AnswerRelevancy: The score is 0.79 because the answer provides some relevant information, but it includes several irrelevant links that do not directly address the question about refusing script rights to a space administrator.

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. For example, the first node discusses 'default right settler policies and conflict resolution,' which does not address the specific question about refusing script rights to an administrator.

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

CustomContextualRelevancy: The retrieval context mentions that rights implied by an admin right will not be overridden by a denial at a lower level, but does not explicitly state that admin rights cannot be refused.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_030_result.json

Overall_score: 0.38666666666666666

Individual_scores:

AnswerRelevancy: 0.9333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.93 because the answer is mostly relevant and helpful, but it includes links to pages on nested pages migration and XWiki syntax, which are not directly related to configuring access and modification settings in XWiki.

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 configuring access controls for team spaces. For example, the first node explains how spaces and pages are created and managed in XWiki, which does not address the input query. Similarly, the second node lists various macros and scripting capabilities, which are also not 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 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.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

File: qa_031_result.json

Overall_score: 0.5264661654135339

Individual_scores:

AnswerRelevancy: 0.7894736842105263

Faithfulness: 1.0

ContextualPrecision: 0.5

ContextualRecall: 0.14285714285714285

CustomContextualRelevancy: 0.2

Reasons:

AnswerRelevancy: The score is 0.79 because while the response provides some useful information, it includes multiple URLs that do not directly address the specific question about authenticating users with access tokens.

Faithfulness: The score is 1.00 because there are no contradictions. Great job!

ContextualPrecision: The score is 0.50 because the first node discusses 'various claims that can be accessed using an access token,' which is not relevant to the authentication of users with access tokens. The second node, which is relevant as it describes the 'Authorized Applications' configuration, is ranked lower. Additionally, nodes ranked 3, 4, and 5 provide information on a security module, cookie protection, and user profile rights, none of which are directly related to the authentication of users with access tokens.

ContextualRecall: The score is 0.14 because the retrieval context only partially aligns with the expected output, specifically in configuring authorized applications (sentence 2), but lacks information on token-based authentication, JWT claims, steps to enable the authenticator, generating JWT tokens, including tokens in requests, and the limitations of the authenticator.

CustomContextualRelevancy: The retrieval context mentions 'Authorized Applications' but lacks details on JWT claims, token generation, and specific configuration steps.

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 content addresses the question about authenticating users with an access token, but the inclusion of URLs that do not directly address the question lowers the relevancy.

Faithfulness: The score is 1.00 because there are no contradictions. Great job maintaining accuracy and faithfulness to 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 discusses 'group synchronization and customization for OpenID Connect,' which is not related to 'using access tokens for authentication.' Similarly, the second node talks about 'protection levels for cookies and security strategies for failed logins,' which is also not relevant.

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.

CustomContextualRelevancy: The retrieval context does not contain any information about token-based authentication for the LLM application (BETA) extension or JWT tokens.

Question_language: de

Expected_answer_language: de

Answer_language: de

File: qa_033_result.json

Overall_score: 0.34666666666666667

Individual_scores:

AnswerRelevancy: 0.7333333333333333

Faithfulness: 1.0

ContextualPrecision: 0

ContextualRecall: 0.0

CustomContextualRelevancy: 0.0

Reasons:

AnswerRelevancy: The score is 0.73 because while the response provides some relevant information, it includes multiple URLs that are not directly related to authenticating users with access tokens, which reduces its 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 0.00 because all nodes in the retrieval context are unrelated to the input. For instance, the first node discusses 'properties of a collection object,' the second node talks about 'access to user profile claims,' and the third node explains 'additional rules for right settler policies,' none of which are 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 about authenticating users with access tokens or JWTs, nor does it mention the specific steps and configurations required for the process.

Question_language: fr

Expected_answer_language: fr

Answer_language: fr

Summarization Results

Model: AI.Models.GPT4o

File: summ_001_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains multiple contradictions, including omitting key actions that can be performed on a page, misrepresenting the purpose of subwikis, and incorrectly specifying the customization of panels. Additionally, it includes extra information not present in the original text and fails to answer several questions that the original text addresses.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.8

Reason: The score is 0.80 because the summary includes extra information not mentioned in the original text and fails to answer a specific question that the original text can address. However, there are no contradictions, indicating a generally accurate representation of the original text.

Score_breakdown:

Alignment: 0.9444444444444444

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains contradictions regarding the limitations for unregistered users and includes extra information not mentioned in the original text. Additionally, it fails to answer specific questions that the original text addresses.

Score_breakdown:

Alignment: 0.7272727272727273

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.7857142857142857

Reason: The score is 0.79 because the summary includes contradictions regarding the upgrade process and preferences, and it fails to address a specific question about migration paths that the original text can answer. However, there is no extra information added that wasn't in the original text.

Score_breakdown:

Alignment: 0.7857142857142857

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.9166666666666666

Reason: The score is 0.92 because the summary is mostly accurate but incorrectly specifies that custom pages outputting RSS formats are created by editing the Main.WebRss page, which is not mentioned in the original text.

Score_breakdown:

Alignment: 0.9166666666666666

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 contains contradictions and extra information not present in the original text, and it fails to answer some questions that the original text addresses.

Score_breakdown:

Alignment: 0.75

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.8

Reason: The score is 0.80 because the summary is generally accurate and does not contain contradictions or unnecessary extra information. However, it misses answering a specific question that the original text addresses.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.8235294117647058

Reason: The score is 0.82 because the summary includes some contradictory information about who can create accounts and adds extra details not mentioned in the original text, such as images illustrating configurations and a link to the next step in the guide.

Score_breakdown:

Alignment: 0.8235294117647058

Coverage: 1.0
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, specifically about XWiki and its references pages.
Score_breakdown:
Alignment: 0.8571428571428571
Coverage: 1.0
Input_language: en
Summary_language: en

File: summ_010_result.json

Score: 0.875
Reason: The score is 0.88 because the summary includes extra information not mentioned in the original text, specifically about the XWiki basic concepts guide. However, there are no contradictions, indicating a generally accurate summary.
Score_breakdown:
Alignment: 0.875
Coverage: 1.0
Input_language: en
Summary_language: en

Model: AI.Models.GPT4o-mini

File: summ_001_result.json

Score: 0.75
Reason: The score is 0.75 because the summary includes contradictory information about panel positioning and extra information about a next section on taking initial steps with XWiki, which are not mentioned in the original text. Additionally, the summary fails to answer a question about exporting a page in PDF format in XWiki 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: 0.8
Reason: The score is 0.80 because the summary includes extra information not mentioned in the original text and fails to answer a specific question that the original text can address.
Score_breakdown:
Alignment: 0.9090909090909091
Coverage: 0.8
Input_language: en
Summary_language: en

File: summ_003_result.json

Score: 0.6

Reason: The score is 0.60 because the summary fails to address specific questions that the original text can answer, indicating that it lacks completeness and detail.

Score_breakdown:

Alignment: 1.0

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.8666666666666667

Reason: The score is 0.87 because the summary includes extra information not present in the original text, such as details about upgrading from Debian Package (.DEB) or Demo Package (.ZIP) and specific commands for updating from Debian Package (.DEB).

Score_breakdown:

Alignment: 0.8666666666666667

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.8

Reason: The score is 0.80 because the summary does not contain any contradictions or extra information, but it fails to answer some questions that the original text can answer.

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 the necessity of logging in and mentions a step not present in the original text. Additionally, it fails to address a specific question about administrative guidance.

Score_breakdown:

Alignment: 0.6

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 accurately reflects the original text without contradictions or unnecessary extra information. However, it fails to address whether there is a 'History'

tab at the bottom of the page to compare versions.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.4

Reason: The score is 0.40 because the summary includes contradictory information about providing visual examples for different configurations and adds extra information about directing users to the next step in creating a basic application. Additionally, the summary fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.8333333333333334

Reason: The score is 0.83 because the summary includes extra information not mentioned in the original text, such as XWiki and references pages. 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: 1.0

Reason: The score is 1.00 because the summary perfectly aligns with the original text, with no contradictions or extra information.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.claude3_5_sonnet

File: summ_001_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains contradictory information about the optional nature and visibility of panels, and it fails to answer specific questions about editing and exporting pages.

Score_breakdown:

Alignment: 0.8
Coverage: 0.6
Input_language: en
Summary_language: en

File: summ_002_result.json

Score: 0.4
Reason: The score is 0.40 because the summary contains significant contradictions with the original text, includes extra information not present in the original text, and fails to answer several questions that the original text addresses.
Score_breakdown:
Alignment: 0.75
Coverage: 0.4
Input_language: en
Summary_language: en

File: summ_003_result.json

Score: 0.8
Reason: The score is 0.80 because the summary is mostly accurate and concise, but it fails to address a specific question that the original text can answer.
Score_breakdown:
Alignment: 1.0
Coverage: 0.8
Input_language: en
Summary_language: en

File: summ_004_result.json

Score: 0.75
Reason: The score is 0.75 because the summary contains some contradictions and extra information not present in the original text, and it fails to answer some questions that the original text can.
Score_breakdown:
Alignment: 0.75
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 extra information not mentioned in the original text and fails to answer a question that the original text can.
Score_breakdown:
Alignment: 0.875
Coverage: 0.8
Input_language: en
Summary_language: en

File: summ_006_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary contains multiple contradictions to the original text, such as the conditions for user login, the documentation of XWiki's programming features, the availability of additional resources, and the inclusion of links to relevant resources.

Score_breakdown:

Alignment: 0.6666666666666666

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 generally accurate and concise, but it fails to address a specific question that the original text can answer.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.3333333333333333

Reason: The score is 0.33 because the summary contains multiple contradictions, extra information not present in the original text, and fails to answer questions that the original text can address.

Score_breakdown:

Alignment: 0.3333333333333333

Coverage: 0.6

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, but it fails to answer some specific questions that the original text can address.

Score_breakdown:

Alignment: 1.0

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 extra information.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.command-r_35B_Q4

File: summ_001_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains contradictions about the basic unit of content in XWiki and the actions that can be performed on a wiki page. Additionally, the summary fails to answer several questions that the original text addresses.

Score_breakdown:

Alignment: 0.6

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains multiple contradictions with the original text and fails to answer some questions that the original text can address.

Score_breakdown:

Alignment: 0.8

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.2

Reason: The score is 0.20 because the summary contains significant contradictions, such as mentioning a 'Preferences' document and 'Access rights' option for terminal pages which are not in the original text. Additionally, it includes extra information about documentation links not found in the original text and fails to answer several specific questions that the original text addresses.

Score_breakdown:

Alignment: 0.7777777777777778

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 mentioned in the original text and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.0

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.8

Reason: The score is 0.80 because the summary is mostly accurate and concise, but it fails to address a specific question that the original text can answer.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.8

Reason: The score is 0.80 because the summary includes a contradiction about the guide providing links to relevant documentation for each role and fails to answer whether there is a dedicated wiki for XWiki Rendering documentation.

Score_breakdown:

Alignment: 0.8

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 a contradiction regarding the features of the 'History' tab, which affects the accuracy. However, there is no extra information or unanswered questions, indicating a generally good summarization.

Score_breakdown:

Alignment: 0.875

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.2

Reason: The score is 0.20 because the summary contains significant contradictions with the original text, includes extra information not mentioned in the original text, and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.5714285714285714

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.8

Reason: The score is 0.80 because the summary fails to answer a specific question that the original text can answer, which is whether it is possible to revert to a previous version of a wiki page.

Score_breakdown:

Alignment: 1.0
Coverage: 0.8
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 not mentioned in the original text, specifically about providing links to explore further details on XWiki's features and extensions.
Score_breakdown:
Alignment: 0.8333333333333334
Coverage: 1.0
Input_language: en
Summary_language: en

Model: AI.Models.gemma2_9B_Q4

File: summ_001_result.json

Score: 0.6666666666666666
Reason: The score is 0.67 because the summary contains contradictory information about the placement of panels and the layout of pages, includes extra information not present in the original text, and fails to address a specific question about exporting a page in PDF format.
Score_breakdown:
Alignment: 0.6666666666666666
Coverage: 0.8
Input_language: en
Summary_language: en

File: summ_002_result.json

Score: 0.2
Reason: The score is 0.20 because the summary includes extra information not mentioned in the original text and fails to answer several questions that the original text can answer.
Score_breakdown:
Alignment: 0.9230769230769231
Coverage: 0.2
Input_language: en
Summary_language: en

File: summ_003_result.json

Score: 0.4
Reason: The score is 0.40 because the summary contains contradictions regarding the use of implicit denial and the configuration of permissions for the Guest user. Additionally, the summary fails to answer several questions that the original text addresses, such as user group creation, permission hierarchy, and the wiki owner's privileges.
Score_breakdown:
Alignment: 0.8571428571428571
Coverage: 0.4

Input_language: en
Summary_language: en

File: summ_004_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains contradictory information regarding the installation process and includes extra details not mentioned in the original text. Additionally, it fails to address certain questions that the original text can answer.

Score_breakdown:

Alignment: 0.8823529411764706

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 contains some contradictions regarding the functionality of the Tags and Watchlist feeds, and it fails to answer whether the default wiki provides a feed for all modified pages.

Score_breakdown:

Alignment: 0.8181818181818182

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.7142857142857143

Reason: The score is 0.71 because the summary contains contradictions regarding the creation of new pages and a specific step called 'Step 1: Creating a page'. Additionally, it fails to answer whether administrators can learn about setting user rights from the guide.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.8571428571428571

Reason: The score is 0.86 because the summary contains a contradiction regarding how to view specific versions, which is a critical detail for understanding the functionality described.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.4

Reason: The score is 0.40 because the summary includes extra information not mentioned in the original text and fails to answer several specific questions that the original text can address.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 0.4

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.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.875

Reason: The score is 0.88 because the summary includes extra information not mentioned in the original text, specifically that XWiki empowers you to create a dynamic and tailored web platform for any purpose.

Score_breakdown:

Alignment: 0.875

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.llama3_1_402b

File: summ_001_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains a contradiction about the optional nature of Panels and fails to address specific questions about editing and exporting pages in XWiki.

Score_breakdown:

Alignment: 0.875

Coverage: 0.6

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 script execution permissions and the presence of a table summarizing permissions. Additionally, the summary fails to answer specific questions about the default status of the comment right and the register right for the non-registered pseudo-user 'XWiki.XWikiGuest'.

Score_breakdown:
Alignment: 0.8823529411764706
Coverage: 0.6
Input_language: en
Summary_language: en

File: summ_003_result.json

Score: 0.6
Reason: The score is 0.60 because the summary contains contradictions regarding implicit denial of rights and the configuration for an open wiki. Additionally, it includes extra information not mentioned in the original text and fails to answer questions about user groups and admin privileges.
Score_breakdown:
Alignment: 0.8235294117647058
Coverage: 0.6
Input_language: en
Summary_language: en

File: summ_004_result.json

Score: 0.6
Reason: The score is 0.60 because the summary contains multiple contradictions with the original text, including incorrect upgrade instructions and redundant claims. Additionally, the summary fails to address specific questions that the original text can answer.
Score_breakdown:
Alignment: 0.8235294117647058
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 contains a contradiction regarding the Watchlist Feature needing to be enabled, and it fails to address whether RSS feeds for modified pages can be received from the Notifications Application.
Score_breakdown:
Alignment: 0.8888888888888888
Coverage: 0.8
Input_language: en
Summary_language: en

File: summ_006_result.json

Score: 0.42857142857142855
Reason: The score is 0.43 because the summary contains multiple contradictions and introduces extra information not present in the original text. Specifically, it incorrectly specifies actions requiring login and mentions links to tutorials and administrative tasks that are not in the original text.
Score_breakdown:
Alignment: 0.42857142857142855
Coverage: 1.0
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 additional information.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.8

Reason: The score is 0.80 because the summary includes contradictions regarding screenshots of rights configurations and checkboxes in the XWiki administration interface, and it fails to address whether admins can create new user accounts in a 'Locked' registration configuration.

Score_breakdown:

Alignment: 0.8181818181818182

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 without any contradictions or extra information.

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 XWiki's flexibility and customizability not mentioned in the original text, and it fails to answer whether there is a list of applications and macros for XWiki available on the extensions wiki.

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: 0.5833333333333334

Reason: The score is 0.58 because the summary contains multiple contradictions with the original text, includes extra information not mentioned in the original text, and fails to answer questions that the original text can address.

Score_breakdown:

Alignment: 0.5833333333333334

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.0

Reason: The score is 0.00 because the summary contains multiple contradictions with the original text and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.625

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.2

Reason: The score is 0.20 because the summary includes extra information not mentioned in the original text and fails to answer several specific questions that the original text can address.

Score_breakdown:

Alignment: 0.9166666666666666

Coverage: 0.2

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, includes extra information not mentioned in the original text, and fails to answer several questions that the original text addresses.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.8

Reason: The score is 0.80 because the summary is mostly accurate and concise, but it fails to address whether it is easy to create new RSS feeds to fit specific needs in XWiki.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en
Summary_language: en

File: summ_006_result.json

Score: 0.5

Reason: The score is 0.50 because the summary contains multiple contradictions to the original text, including the scope of the guide, missing links in the Developers section, and an incorrect next step in the guide.

Score_breakdown:

Alignment: 0.5

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.6

Reason: The score is 0.60 because the summary contains contradictory information regarding how to view and rollback versions. Additionally, it fails to answer specific questions about content saving and the presence of a 'History' tab.

Score_breakdown:

Alignment: 0.8571428571428571

Coverage: 0.6

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary contains multiple contradictions with the original text, including incorrect details about the configurations for view and edition rights, registration settings, and the actions allowed in different configurations. Additionally, the summary fails to answer a question that the original text can address.

Score_breakdown:

Alignment: 0.6666666666666666

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 includes extra information not mentioned in the original text and fails to answer a question that the original text can answer.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.8

Reason: The score is 0.80 because the summary includes extra information not mentioned in the original text and fails to answer a question that the original text can address.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 0.8

Input_language: en

Summary_language: en

Model: AI.Models.mistral-nemo_12b_Q4

File: summ_001_result.json

Score: 0.8

Reason: The score is 0.80 because the summary contains some incorrect information regarding the Footer Area and user options for column display and width. Additionally, the summary fails to address whether a user can export a page in PDF format in XWiki.

Score_breakdown:

Alignment: 0.8181818181818182

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.0

Reason: The score is 0.00 because the summary contains multiple contradictions to the original text, includes extra information not mentioned in the original text, and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.6923076923076923

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.4

Reason: The score is 0.40 because the summary includes incorrect information about setting access rights for terminal pages, adds details about documentation not mentioned in the original text, and fails to address several questions that the original text can answer.

Score_breakdown:

Alignment: 0.6

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.25

Reason: The score is 0.25 because the summary contains multiple contradictions with the original text, includes extra information not mentioned in the original text, and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.25

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 1.0

Reason: The score is 1.00 because the summary accurately reflects the original text without any contradictions or additional information.

Score_breakdown:

Alignment: 1.0

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 1.0

Reason: The score is 1.00 because the summary accurately reflects the original text without any contradictions or extraneous information.

Score_breakdown:

Alignment: 1.0

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 contains a contradiction regarding the 'More Actions' button's functionality, which affects the accuracy. However, there are no extra pieces of information or unanswered questions, indicating a generally good summary.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.8

Reason: The score is 0.80 because the summary contains several contradictions regarding the settings and permissions described in the original text. Additionally, the summary fails to answer a specific question about accessing the global rights Administration in XWiki.

Score_breakdown:

Alignment: 0.8235294117647058

Coverage: 0.8

Input_language: en
Summary_language: en

File: summ_009_result.json

Score: 0.5454545454545454

Reason: The score is 0.55 because the summary contradicts the original text regarding the advantages of wikis, includes extra information not present in the original text, and fails to answer a question that the original text addresses.

Score_breakdown:

Alignment: 0.5454545454545454

Coverage: 0.8

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 the next section in the guide covering XWiki basic concepts, which is not mentioned in the original text.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.mistral2_large

File: summ_001_result.json

Score: 0.8947368421052632

Reason: The score is 0.89 because the summary contains minor contradictions about the placement of panels and includes extra information about the Content area that is not specified in the original text.

Score_breakdown:

Alignment: 0.8947368421052632

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.9166666666666666

Reason: The score is 0.92 because the summary includes extra information about a table with additional information that is not mentioned in the original text.

Score_breakdown:

Alignment: 0.9166666666666666

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.8333333333333334

Reason: The score is 0.83 because the summary contains some contradictions regarding access rights settings and includes extra information not mentioned in the original text. However, it still captures the main points accurately.

Score_breakdown:

Alignment: 0.8333333333333334

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.76

Reason: The score is 0.76 because the summary introduces several contradictions, such as dividing the guide into parts and mentioning specific methods for upgrading that are not present in the original text. However, there is no extra information added.

Score_breakdown:

Alignment: 0.76

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 accurately reflects the original text without any contradictions or additional information.

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 contains contradictions regarding user login requirements and the next steps in the guide. Additionally, the summary fails to answer specific questions that the original text can address.

Score_breakdown:

Alignment: 0.6666666666666666

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 changing the logo and the panels that is not mentioned in the original text. However, there are no contradictions, indicating a generally accurate summary.

Score_breakdown:

Alignment: 0.875
Coverage: 1.0
Input_language: en
Summary_language: en

File: summ_008_result.json

Score: 0.4
Reason: The score is 0.40 because the summary includes information that contradicts the original text regarding rights management in XWiki. Additionally, the summary fails to address several questions that the original text can answer, indicating significant omissions and inaccuracies.
Score_breakdown:
Alignment: 0.8571428571428571
Coverage: 0.4
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.
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.
Score_breakdown:
Alignment: 1.0
Coverage: 1.0
Input_language: en
Summary_language: en

Model: AI.Models.mixtral-8x22b

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 extra information. However, it fails to address whether a user can export a page in PDF format.
Score_breakdown:
Alignment: 1.0
Coverage: 0.8
Input_language: en
Summary_language: en

File: summ_002_result.json

Score: 0.8

Reason: The score is 0.80 because the summary contains some contradictions regarding the permissions that can be granted or revoked and omits specific details about a table summarizing rights. Additionally, the summary fails to answer a question about the default status of the comment right.

Score_breakdown:

Alignment: 0.8666666666666667

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.8461538461538461

Reason: The score is 0.85 because the summary contains a contradiction regarding the configuration of permissions for an open wiki and includes extra information about documentation for sub-wiki access rights that is not mentioned in the original text.

Score_breakdown:

Alignment: 0.8461538461538461

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.8

Reason: The score is 0.80 because the summary contains multiple pieces of contradicting information and extra details not mentioned in the original text. Additionally, it fails to answer a specific question that the original text addresses.

Score_breakdown:

Alignment: 0.8125

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.9166666666666666

Reason: The score is 0.92 because the summary contains a minor contradiction regarding how RSS feeds can be enabled, but otherwise it is accurate and concise.

Score_breakdown:

Alignment: 0.9166666666666666

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.8181818181818182

Reason: The score is 0.82 because the summary includes minor contradictions about the purpose of pre-installed applications and adds extra steps not mentioned in the original text. However, it still captures the main points well.

Score_breakdown:

Alignment: 0.8181818181818182

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 fails to address a specific question that the original text can answer.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_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_009_result.json

Score: 0.8

Reason: The score is 0.80 because the summary is mostly accurate and does not contain any contradictions or extra information. However, it fails to address a specific question that the original text can answer, which slightly reduces its completeness.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.8

Reason: The score is 0.80 because the summary fails to answer a specific question that the original text can address, indicating a gap in the provided information.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en
Summary_language: en

Model: AI.Models.phi3_medium-128k_14b_Q4

File: summ_001_result.json

Score: 0.6363636363636364

Reason: The score is 0.64 because the summary includes multiple contradictions and extra information not present in the original text, which affects its accuracy and reliability.

Score_breakdown:

Alignment: 0.6363636363636364

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 significant contradictions regarding the default status of the view right, which is misrepresented as 'deny' or 'no setting' instead of 'ALLOWED' at both the Page and Wiki levels. Additionally, the summary fails to answer several key questions that the original text addresses, such as the default status of the comment right and details about delete, administration, and programming rights.

Score_breakdown:

Alignment: 0.7142857142857143

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.2

Reason: The score is 0.20 because the summary contains significant contradictions with the original text, introduces extra information not mentioned in the original text, and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.6923076923076923

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.2

Reason: The score is 0.20 because the summary contains significant contradictions and a large amount of extra information not present in the original text. Additionally, there are several questions that the original text can answer but the summary cannot.

Score_breakdown:

Alignment: 0.42857142857142855

Coverage: 0.2

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.5555555555555556

Reason: The score is 0.56 because the summary includes contradictory information about the Watchlist feed generation and adds extra details not mentioned in the original text, such as XWiki being open-source and specifics about RSS feed features.

Score_breakdown:

Alignment: 0.5555555555555556

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.7777777777777778

Reason: The score is 0.78 because the summary incorrectly assigns tasks meant for administrators to developers and includes extra information about a section ending with a direct link to 'Step 1: Creating a page' that is not mentioned in the original text.

Score_breakdown:

Alignment: 0.7777777777777778

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 contains a contradiction regarding the navigation method and includes extra information about a detailed guide link that is not mentioned in the original text.

Score_breakdown:

Alignment: 0.75

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains multiple pieces of contradicting information regarding user rights, protection levels, and the steps involved in creating an application. Additionally, it fails to address several questions that the original text can answer, indicating significant gaps in coverage and accuracy.

Score_breakdown:

Alignment: 0.625

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.8

Reason: The score is 0.80 because the summary is generally accurate and concise, but it fails to address some specific questions that the original text can answer, such as whether a wiki allows you to modify a page by clicking an 'edit' button.

Score_breakdown:

Alignment: 1.0

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.875

Reason: The score is 0.88 because the summary includes extra information not mentioned in the original text, specifically about a user guide chapter on 'XWiki Basic Concepts.' However, there are no contradictions, which indicates a generally high-quality summary.

Score_breakdown:

Alignment: 0.875

Coverage: 1.0

Input_language: en

Summary_language: en

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 key questions that the original text addresses, indicating it lacks essential information.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.0

Reason: The score is 0.00 because the summary contains multiple contradictions with the original text, includes extra information not mentioned in the original text, and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.0

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 contains multiple contradictions with the original text, includes extra information not mentioned in the original text, and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.3333333333333333

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.2

Reason: The score is 0.20 because the summary contains multiple 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.

Score_breakdown:

Alignment: 0.5263157894736842

Coverage: 0.2

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 multiple key questions that the original text can answer, indicating a significant lack of coverage and completeness.

Score_breakdown:

Alignment: 0

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.75

Reason: The score is 0.75 because the summary omits the user role and incorrectly states the starting tasks, and it includes extra information about managing users for access control that is not in the original text.

Score_breakdown:

Alignment: 0.75

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_007_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 several questions that the original text can answer.

Score_breakdown:

Alignment: 0.0

Coverage: 0.0

Input_language: en
Summary_language: en

File: summ_008_result.json

Score: 0.375

Reason: The score is 0.38 because the summary includes multiple contradictions and extra information not present in the original text, and it fails to answer a question that the original text can.

Score_breakdown:

Alignment: 0.375

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.7692307692307693

Reason: The score is 0.77 because the summary includes some contradictions and extra information not present in the original text. Additionally, there are questions that the original text can answer but are not addressed in the summary. Despite these issues, the summary still captures a significant portion of the original text's content.

Score_breakdown:

Alignment: 0.7692307692307693

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 mentioned in the original text, such as specific deployment environments, responsive design features of XWiki, and details about the XWiki User Guide.

Score_breakdown:

Alignment: 0.75

Coverage: 1.0

Input_language: en

Summary_language: en

Model: AI.Models.qwen2_7b_Q4

File: summ_001_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary contains contradictory information about the tabbed area at the bottom of the page and includes extra information not mentioned in the original text about XWiki being a collaborative platform and its design for facilitating collaboration and knowledge management.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_002_result.json

Score: 0.0

Reason: The score is 0.00 because the summary contains multiple contradictions with the original text, includes extra information not mentioned in the original text, and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.4117647058823529

Coverage: 0.0

Input_language: en

Summary_language: en

File: summ_003_result.json

Score: 0.4

Reason: The score is 0.40 because the summary contains contradictory information about where Global Users/Groups are defined, includes extra information not mentioned in the original text, and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.75

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_004_result.json

Score: 0.4

Reason: The score is 0.40 because the summary includes extra information not mentioned in the original text and fails to answer several questions that the original text can address.

Score_breakdown:

Alignment: 0.8461538461538461

Coverage: 0.4

Input_language: en

Summary_language: en

File: summ_005_result.json

Score: 0.75

Reason: The score is 0.75 because the summary includes contradictions regarding the use of a template from Main.WebRss and the customization of RSS content for specific applications like Notifications and Search, which were not mentioned in the original text.

Score_breakdown:

Alignment: 0.75

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_006_result.json

Score: 0.6666666666666666

Reason: The score is 0.67 because the summary contains some contradictions and extra information not present in the original text. Specifically, it incorrectly states the guide is divided into three parts and misrepresents the contents of the Developer's Guide. Additionally, it adds information about the next step being creating a page, which is not mentioned in the original text.

Score_breakdown:

Alignment: 0.6666666666666666

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_007_result.json

Score: 0.625

Reason: The score is 0.62 because the summary contains several contradictions with the original text, such as the location of the 'History' tab and the function of the 'Compare selected versions' button. Additionally, the summary fails to mention certain steps and features described in the original text. There is also a question that the original text can answer but the summary cannot.

Score_breakdown:

Alignment: 0.625

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_008_result.json

Score: 0.9

Reason: The score is 0.90 because the summary includes extra information not mentioned in the original text, such as specific configurations demonstrated in images and details about permissions and protection strength levels.

Score_breakdown:

Alignment: 0.9

Coverage: 1.0

Input_language: en

Summary_language: en

File: summ_009_result.json

Score: 0.7

Reason: The score is 0.70 because the summary includes extra information not mentioned in the original text and fails to answer a question that the original text can answer.

Score_breakdown:

Alignment: 0.7

Coverage: 0.8

Input_language: en

Summary_language: en

File: summ_010_result.json

Score: 0.8571428571428571

Reason: The score is 0.86 because the summary includes extra information about XWiki's flexibility that is not mentioned in the original text. However, there are no contradictions, indicating a generally accurate summary.

Score_breakdown:
Alignment: 0.8571428571428571
Coverage: 1.0
Input_language: en
Summary_language: en

Text_generation Results

Model: AI.Models.GPT4o

File: text_gen_001_result.json

Score: 0.8
Reason: The Actual Output addresses the topic and includes all key information (name, date of birth, address) from the Input. However, the format does not match the Expected Output JSON structure.
Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9
Reason: The Actual Output addresses the Input prompt and aligns closely with the Expected Output. It is coherent and maintains consistency, with only a slight variation in word choice ('frolicked' vs 'enjoyed exploring'). The language is grammatically correct and natural.
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 relevant, coherent, and fluent. It effectively adapts the text for a younger audience by comparing photosynthesis to a magic trick, which is engaging for children. However, it slightly deviates from the Expected Output by not explicitly mentioning 'a special process called photosynthesis.'
Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_004_result.json

Score: 1.0
Reason: The Actual Output directly addresses the Input prompt, covering all necessary aspects such as noise-cancelling, 20-hour battery life, and comfortable fit for tech-savvy professionals. It aligns well with the Expected Output, presenting key points and information in a coherent and fluent manner.
Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_005_result.json

Score: 0.9

Reason: The Actual Output directly addresses the topic of a space adventure and covers key points such as interstellar travel, exploration of a new galaxy, and interaction with alien beings. It is coherent and well-organized, with a logical flow and smooth language. However, it slightly exceeds the Expected Output by including a broader narrative beyond just 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 directly addresses the Input prompt by writing a poem about the beauty of nature. It stays on topic throughout, logically flows from one stanza to the next, and maintains coherence. The grammar, syntax, and overall readability are excellent. The Actual Output aligns perfectly with the Expected Output in terms of relevance, coherence, and fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by providing a detailed news article about a recent scientific discovery. It aligns with the Expected Output by summarizing a groundbreaking finding. The article is coherent, logically flows, and is free of grammatical errors, making it fluent and easy to read.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by providing a comprehensive travel guide for Kyoto, Japan. It is coherent and aligns well with the Expected Output, covering top attractions, accommodations, and activities. The fluency is excellent with proper grammar, syntax, and readability.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output addresses the Input prompt and is relevant to the Expected Output. It is coherent and logically structured. However, it does not match the format and specific key names of the Expected Output, and it combines the start and end dates into one field.

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 Input prompt and stays on topic, covering key points such as regulation of physiological processes and cognitive functions. However, it misses some details like 'sophisticated,' 'coordinating,' and 'subjective experiences' from 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: 0.9

Reason: The Actual Output correctly extracts the name, date of birth, and address from the Input and is coherent and fluent, but it does not match the Expected Output format.

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 prompt and is relevant to the Expected Output. It is coherent and logically flows, maintaining consistency with the Expected Output. The fluency is high with correct grammar and natural language use. The slight difference in wording does not significantly impact the overall quality.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively addresses the topic of photosynthesis for a younger audience, using relatable metaphors like 'magic trick' and 'superpower.' It covers the key points of using sunlight to make food and grow, similar to the Expected Output. The text is coherent and fluent, though slightly more playful than the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by describing the Wireless Bluetooth Headphones, including features like noise-cancelling, 20-hour battery life, and comfortable fit, targeting tech-savvy professionals. It maintains coherence and logical flow similar to the Expected Output. The grammar, syntax, and readability are fluent and meet the quality of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.7

Reason: The Actual Output is a creative story about a space adventure, addressing the Input prompt well. However, it diverges from the Expected Output by focusing on a nebula rather than a new planet. The story is coherent, fluent, and engaging, but it misses the specific detail of exploring a new planet.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by describing the beauty of nature and aligns with the Expected Output. It is coherent, flowing logically from one stanza to the next, and maintains consistency throughout. The poem is fluent, with correct grammar, natural language use, and high readability.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by generating a news article about a recent scientific discovery. It is coherent with the Expected Output, providing a detailed and logical summary of the groundbreaking finding. The article is fluent, with proper grammar, syntax, 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 directly addresses the topic by providing a detailed travel guide for Kyoto, Japan. It covers all key points such as top attractions, cultural experiences, transportation, and tips for travelers, which align with the Expected Output. The guide is well-organized, coherent, and free from grammatical errors, ensuring a smooth reading experience.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.7

Reason: The Actual Output addresses the Input prompt and includes all the necessary details, but the format and some specific elements differ from the Expected Output. The dates are not split into start_date and end_date, and the target_audience is not explicitly mentioned.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output directly addresses the topic and closely matches the Expected Output, covering all key points and details. It maintains coherence and logical flow, with minor differences in word choice that do not affect the technical accuracy or readability.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.claude3_5_sonnet

File: text_gen_001_result.json

Score: 0.9

Reason: The Actual Output correctly extracts and presents the key information (name, date of birth, address) from the Input text. However, it does not match the Expected Output format exactly, as it uses a different structure.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input prompt and closely aligns with the Expected Output. It is coherent, grammatically correct, and uses natural language. The main difference is the choice of verbs 'frolicked' vs. 'enjoyed exploring', which slightly affects completeness but not 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 directly addresses the Input prompt by explaining photosynthesis in a way that is engaging for a younger audience. It is coherent and logically flows with the Expected Output, adding a bit more detail and analogy to make it relatable. The fluency is high, with proper grammar and readability.

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 topic and covers all key points from the Expected Output, including noise-cancelling, 20-hour battery life, and comfortable fit. It is coherent and fluent with logical flow and good readability. Minor differences in phrasing and additional details slightly differentiate it

from 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 directly addresses the prompt by providing a creative space adventure story. However, it diverges from the Expected Output, which specifies a story about exploring a new planet. The Actual Output involves a wormhole and an alien megastructure instead. The story is coherent and fluent with good grammar and readability.

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 clearly addresses the Input prompt by describing the beauty of nature, matching the Expected Output of describing the wonders of the natural world.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by generating a news article about a recent scientific discovery. It aligns with the Expected Output by summarizing a groundbreaking finding in Alzheimer's research. The article is coherent, logically flowing from the introduction of the drug to its trial results and future implications. It is fluent, with proper grammar, syntax, and readability matching the quality of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by providing a comprehensive travel guide for Paris, France, a popular tourist destination. It aligns with the Expected Output by highlighting top attractions, accommodations, and activities. The guide is coherent, logically flows, and maintains consistency throughout. It is also grammatically correct and uses natural language. There are no significant deviations or missing elements compared to 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 addresses the Input prompt and covers most key points such as event name, dates, venue, and address. However, it combines the start and end dates into a single date range and includes an event description instead of specifically listing the target audience.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input prompt and matches the context of the Expected Output. It is coherent, fluent, and maintains a consistent narrative. Key points such as the brain's role in physiological functions, sensory processing, and cognitive functions are accurately reflected. However, it includes additional details like the number of neurons and synaptic connections, which, while informative, slightly diverge from the Expected Output's focus.

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 addresses the topic and extracts the key information accurately. It matches the Expected Output in terms of content but uses a different format. The text is coherent and fluent.

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 covers all aspects. It is coherent and fluent. However, it deviates slightly from the Expected Output in structure and some details.

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 topic of photosynthesis and adapts it well for a younger audience. It covers all key points from the Expected Output, explaining how plants use sunlight to make energy. The explanation is coherent and flows logically, with natural language use and grammatical correctness. However, it includes additional details like 'delicious energy' which, while engaging, are not strictly necessary.

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 topic by describing the wireless Bluetooth headphones and their features. It covers all key points such as noise-cancelling, 20-hour battery life, and comfortable fit, similar to the Expected Output. The ideas flow logically and are well-organized, with good grammar and readability. The only minor difference is the slightly different wording and structure.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 1.0

Reason: The Actual Output directly addresses the topic of a space adventure and aligns with the Expected Output of a group of astronauts exploring a new planet. It covers all key points with a coherent and fluent narrative, detailing the journey, discovery, and return.

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 prompt by describing the beauty of nature and aligns with the Expected Output of a poem about the natural world. The poem is coherent, logically flowing from one stanza to the next, and maintains consistency in its theme. The fluency is high, with correct grammar, natural phrasing, and readability.

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, addressing the recent scientific discovery of a 'star quiz' or microwave burst. It is coherent, follows a logical structure, and maintains consistency. The article is fluent, with correct grammar and appropriate vocabulary. It aligns well with the Expected Output by providing an informative summary of the scientific finding. Minor inaccuracies, such as the term 'star quiz' instead of 'star quirk,' slightly affect the score.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output directly addresses the topic by providing a comprehensive travel guide for Paris, covering key attractions, accommodations, and activities. It is detailed, coherent, and fluent, matching the Expected Output perfectly.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output addresses the Input prompt and is relevant to the Expected Output. It is coherent and fluent, but it lacks the specific format and some details such as 'start_date', 'end_date', 'venue', 'address', and 'target_audience' as specified 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 prompt and covers key points such as physiological functions, sensory information, and cognitive processes. However, it lacks specific terms like 'biological substrate' and 'subjective experiences' found in the Expected Output. The coherence and fluency are strong, with logical flow and good readability.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.gemma2_9B_Q4

File: text_gen_001_result.json

Score: 0.9

Reason: The Actual Output addresses the topic by extracting the key information correctly and covers all key points from the Expected Output. It is coherent and fluent, but the format differs slightly from 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 addresses the Input prompt and aligns with the Expected Output, though it uses different phrasing. It is coherent, grammatically correct, and stays on topic, but the focus on 'basking in the sunshine' slightly deviates from 'exploring 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 addresses the topic of photosynthesis and adapts it well for a younger audience. It aligns with the Expected Output by explaining that plants use sunlight to make food. The explanation is coherent and uses natural language appropriate for ages 8-10. However, it includes additional imagery 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 directly addresses the Input prompt and matches the context of the Expected Output. It highlights the key features: noise-cancelling, 20-hour battery life, and comfortable fit, targeting tech-savvy professionals. The coherence and fluency are maintained with correct grammar and readability. 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: 0.7

Reason: The Actual Output is a creative space adventure that addresses the topic posed by the Input. However, it does not match the Expected Output, which specifies a group of astronauts exploring a new planet. The story focuses on a single character, Lyra, and her journey to find a mythical planet, which deviates from the group dynamic and exploration theme. The narrative is coherent and fluent, with proper grammar and punctuation.

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 prompt by providing a poem about the beauty of nature. It matches the context of the Expected Output by describing the wonders of the natural world. The poem is coherent, with each stanza logically flowing from one to the next, and it is fluent, with correct grammar, natural language use, and high readability.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by generating a news article about a recent scientific discovery. It aligns with the Expected Output by summarizing a groundbreaking scientific finding. The article is coherent, with a logical flow and consistency in ideas, and it is fluent, with grammatically correct and natural language.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by providing a detailed travel guide for Rome. It covers key points and themes such as top attractions, food, and travel tips, aligning well with the Expected Output. The guide is coherent, logically structured, and written in fluent, grammatically correct language.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output is relevant to the Input and captures key details such as event name, dates, and location. However, it misses the target audience and does not separate start and end dates 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 addresses the Input prompt and covers all required aspects, including the complexity of the brain and its functions. It includes key points from the Expected Output, such as physiological processes, sensory perception, and cognitive functions. The coherence and fluency are strong, with logical flow and good readability. However, it slightly diverges by adding a sentence on the subjective experiences of consciousness, which, while relevant, was not in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.llama3_1_402b

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output correctly extracts the key information (name, date of birth, address) from the Input text and presents it clearly. However, it does not match the Expected Output format, which is a JSON object.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.8

Reason: The Actual Output addresses the Input prompt and includes all keywords. It is more detailed than the Expected Output but still coherent and fluent. However, it slightly deviates from the simplicity of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output effectively addresses the topic of photosynthesis for a younger audience, covering the key points of plants making their own food using sunlight. It is coherent and fluent, with a logical flow and natural language. However, it introduces additional elements like 'yummy, sweet drink' which are not present in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 1.0

Reason: The Actual Output directly addresses the topic by describing the wireless Bluetooth headphones for tech-savvy professionals. It covers all key points from the Expected Output including noise-cancelling technology, 20-hour battery life, and comfortable fit. The description is coherent, logically flows, and is grammatically accurate.

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 story about astronauts exploring a new planet, aligning well with the Expected Output and addressing the Input prompt directly.

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 prompt by providing a poem about the beauty of nature. It covers all aspects mentioned, including various elements of nature like mountains, sun, trees, rivers, flowers, stars, and the moon. The poem is coherent, flows logically, and is grammatically correct. It aligns well with the Expected Output, 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 prompt by providing a detailed news article about a recent scientific discovery. The ideas are logically connected, and the text is coherent and easy to follow. The language is fluent with correct grammar and smooth flow. The key points and overall message align with the Expected Output, summarizing a groundbreaking scientific finding.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by providing a detailed travel guide for Santorini, Greece, which aligns with the Expected Output. It is coherent, logically flowing from one section to the next, and is fluent with correct grammar and natural language use. The guide includes top attractions, accommodations, activities, and practical tips, meeting the anticipated quality.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.7

Reason: The Actual Output addresses the Input prompt and extracts key event details such as the event name, date, and location. However, it misses specific details like the start and end dates, venue, address, and target audience mentioned in the Expected Output. The information is coherent and fluent but lacks completeness.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output directly addresses the topic and uses technical terminology appropriately. It covers all key points from the Expected Output, such as the brain's role in physiological processes, sensory information processing, and cognitive functions. The ideas flow logically and the text is coherent and fluent. Minor differences in wording do not detract from the overall accuracy and clarity.

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 addresses the Input prompt and extracts the correct key information. However, the format differs from the Expected Output, which uses a JSON structure instead of a list format.

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 includes all keywords. It is coherent and grammatically correct. However, it is more elaborate than the Expected Output and introduces additional details not present in 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 addresses the topic of photosynthesis and adapts it well for a younger audience. It covers key points such as plants using sunlight to make food and the process being called photosynthesis. The ideas flow logically and are well-organized, with appropriate vocabulary and natural phrasing. However, 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 is relevant to the Input prompt and aligns well with the Expected Output in terms of addressing the product features and target audience. It is coherent and logically structured, with a clear focus on the benefits for tech-savvy professionals. The fluency is high, with proper grammar and syntax. Minor differences in phrasing and emphasis prevent a perfect score.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.9

Reason: The Actual Output directly addresses the topic of a space adventure, covering key points of exploration and discovery. It includes detailed characters and a coherent plot with logical flow. The language is fluent and grammatically correct. However, it slightly diverges from the Expected Output by focusing on a mission beyond the galaxy rather than just 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 directly addresses the topic of the beauty of nature as posed in the Input. It covers all key points expected in a poem about the natural world, including descriptions of the moon, trees, sun, flowers, river, creatures, and stars. The poem is coherent with a logical flow and structure, and it is fluent with correct grammar and natural language use.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by providing a detailed news article about a recent scientific discovery. It aligns with the Expected Output by summarizing a

groundbreaking finding in quantum computing. The article is coherent with a logical flow and consistent ideas, and it is fluent with proper grammar, syntax, 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 directly addresses the topic by providing a comprehensive travel guide for Paris. It covers key points such as top attractions, accommodations, and practical information, aligning well with the Expected Output. The guide is coherent, logically structured, and fluent with correct grammar and natural language use.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output is relevant to the Input and coherent, with correct grammar and natural language use. However, it 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: 0.8

Reason: The Actual Output addresses the topic and covers key points such as bodily functions and sensory information, similar to the Expected Output. However, it lacks mention of 'consciousness' and 'subjective experiences.' The coherence and fluency are strong, with clear logical flow and correct grammar.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.mistral-nemo_12b_Q4

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output addresses the topic and covers all key points from the Expected Output, but the format differs.

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 topic and includes all key points from the Expected Output but has a slightly different structure. It is coherent 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 addresses the topic of photosynthesis and explains it in a way that is engaging for a younger audience. It aligns well with the Expected Output by mentioning that plants use sunlight to make food and grow. The explanation is coherent and logically flows from one idea to the next. The language is natural, grammatically correct, and easy to read. However, it includes some additional details like chloroplasts and glucose that were not in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input and Expected Output, maintaining coherence and fluency. It effectively highlights the product's features and target audience, though it slightly differs in phrasing and structure.

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 story about a space adventure, addressing the Input. However, it diverges from the Expected Output as it focuses on musical themes and time travel rather than exploring a new planet. The story is coherent, grammatically correct, and uses natural language.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output addresses the main point of writing a poem about the beauty of nature. It covers all key elements and details expected, including descriptions of forests, rivers, mountains, fields, waterfalls, and creatures. The ideas are logically connected and flow smoothly, with coherent and natural sentence structure, appropriate vocabulary, and grammatical correctness.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output is highly relevant to the Input prompt, providing a detailed and coherent news article about a recent scientific discovery. It logically flows from the headline to the conclusion, maintaining grammatical correctness and natural language usage throughout. The key elements and intent of the Expected Output are fully met, summarizing the groundbreaking finding effectively.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output addresses the main points of creating a travel guide, covering top attractions, accommodations, and activities in Santorini. It includes detailed sections on the best time to visit, getting there, getting around, top attractions, tips, a sample itinerary, and local cuisine. The content is coherent, logically connected, and fluent with appropriate word choice and grammatical correctness.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.7

Reason: The Actual Output is relevant and coherent, addressing the main points of the Input prompt. However, it lacks details such as the start and end dates separately, the target audience, and the address is split into multiple lines instead of a single string.

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 Input prompt and is relevant. It covers key points such as brain structure and functions, similar to the Expected Output. However, it uses terms like 'neocortex' and 'cerebrum' which are not in the Expected Output, and the phrase 'harboring the cognitive and consciousness aspects' is less precise than 'biological substrate for cognitive functions, consciousness, and subjective experiences'. The coherence and fluency are maintained.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.mistral2_large

File: text_gen_001_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input prompt and covers similar topics and details as the Expected Output. The information is coherent and logically presented. However, the format differs slightly from 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 and maintains the key points and themes of the Expected Output. However, there is a slight difference in wording and tense, which affects the coherence and fluency slightly.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output simplifies the concept of photosynthesis effectively for a younger audience and aligns well with the Expected Output. It maintains coherence and fluency, but it slightly differs in wording by not explicitly mentioning 'energy' and instead using 'food'.

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 and aligns well with the Expected Output in terms of key points and themes. It mentions noise-cancelling, 20-hour battery life, and comfortable fit, targeting tech-savvy professionals. The logical flow and readability are excellent, with only minor differences in phrasing.

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, addressing the Input request. However, it diverges from the Expected Output which specifically mentions astronauts exploring a new planet. The story is coherent, fluent, and well-written, but it does not fully align with the key points of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is highly relevant to the Input prompt, addressing the beauty of nature comprehensively. It meets the key points of the Expected Output by vividly describing the wonders of the natural world. The ideas are logically connected and flow smoothly, and the poem is free of grammatical errors, reading naturally.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by providing a detailed news article about a recent scientific discovery. The article is coherent, with a logical flow and consistent ideas. It is fluent, with proper grammar, syntax, and readability. The content meets the Expected Output by summarizing a groundbreaking scientific finding.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output directly addresses the Input prompt by providing a comprehensive travel guide for Paris, a popular tourist destination. It is logically structured with well-connected ideas, covering the best time to visit, getting around, top attractions, food and drink, shopping, day trips, tips for travelers, and emergency numbers. The fluency is on par with the Expected Output, reading smoothly and naturally. It meets the specific requirements by highlighting top attractions, accommodations, and activities.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output addresses the topic and includes most key details such as event name, dates, venue, address, and activities. However, it does not separate the start and end dates and includes 'Audience' instead of 'target_audience'. The coherence and fluency are well-maintained.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output directly addresses the topic and covers all key points from the Expected Output, including regulation of physiological functions, processing sensory data, and serving as the substrate for cognition and consciousness. It is coherent and fluent, with minor differences in wording but no significant deviations in meaning.

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 addresses the topic and covers all key points (name, date of birth, address) but does not match the JSON format of the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input prompt and aligns with the Expected Output in terms of relevance. It is coherent and logically flows from one idea to the next. The only minor difference is the use of 'loves to frolic' instead of 'enjoyed exploring,' but both convey a similar meaning and maintain fluency.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input prompt by simplifying the concept of photosynthesis for a younger audience. It covers all key points from the Expected Output, such as plants making their own food using sunlight and turning it into energy. The explanation is coherent, logically organized, and uses age-appropriate language. Minor deviation in details like mentioning chloroplasts and sugar adds educational value without overwhelming 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 directly addresses the Input prompt and aligns closely with the Expected Output, covering key points like noise-cancelling, 20-hour battery life, and comfortable fit for tech-savvy professionals. The coherence and fluency are strong, with minor differences in wording.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input prompt and covers all required aspects by telling a creative story about a space adventure. It includes key points such as the exploration of new planets and unexpected discoveries, aligning well with the Expected Output. The story is coherent with a logical flow and consistent ideas, and the language used is fluent and grammatically correct. However, it slightly deviates by including additional elements like the ancient civilization, which were not specified in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 1.0

Reason: The Actual Output is highly relevant to the Input prompt, directly addressing the beauty of nature. The poem is coherent with well-organized ideas and logical flow. It is also fluent, with proper grammar and syntax. The content, structure, and quality align perfectly with the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the topic of a recent scientific discovery, specifically a potential cure for Type 1 Diabetes. It covers all key points and details expected in an informative article, including the discovery, the research team, and the clinical trials. The article is coherent with a logical flow and consistent ideas, and it is fluent with correct 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 directly addresses the topic of creating a travel guide for Kyoto, Japan. It covers all key points and details such as top attractions, food, accommodations, transportation, and tips, matching the Expected Output. The guide is coherent, logically structured, and fluent with correct grammar and syntax.

Prompt_language: ro

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.8

Reason: The Actual Output is relevant to the Input and covers key points such as event name, dates, and location. However, it does not separate the start and end dates, and it lacks the 'target_audience' field.

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 prompt, addressing the complexity and functions of the brain with technical language. It is coherent and fluent, with logical flow and grammatical correctness. It includes additional technical details and research methods, which are aligned with the Expected Output's emphasis on complexity and cognitive functions.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.phi3_medium-128k_14b_Q4

File: text_gen_001_result.json

Score: 0.8

Reason: The Actual Output addresses the Input prompt by extracting the key information correctly. However, it does not follow the JSON structure 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 addresses the Input prompt and matches the context of the Expected Output. It is coherent, grammatically correct, and uses natural language. While it uses different phrasing, the main ideas and details align closely with the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.8

Reason: The Actual Output addresses the Input prompt and matches the context of the Expected Output by explaining photosynthesis in a way that is accessible to a younger audience. It is coherent and flows logically, using relatable analogies. However, it is more detailed and longer than the Expected Output, which is simpler and more concise.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.8

Reason: The Actual Output addresses the Input prompt and aligns well with the Expected Output, covering key points like noise-cancelling, 20-hour battery life, and comfortable fit for tech-savvy professionals. It is coherent and fluent, though there is a minor typo ('cutting- vice').

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 Input prompt by providing a creative space adventure story and aligns well with the Expected Output by focusing on astronauts exploring new frontiers. However, it deviates slightly by including an asteroid mining subplot, which adds complexity but slightly strays from the core theme of exploring a new planet. The story is coherent, fluent, and well-written.

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 topic of the beauty of nature, covering all key points and details expected in a poem about the wonders of the natural world. The poem is coherent, with a logical flow and consistency in the text, and it is fluent, with grammatical correctness and natural language use.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 0.8

Reason: The Actual Output is relevant to the Input prompt and addresses the main topic of a recent scientific discovery. It is coherent and logically flows, maintaining a consistent narrative. The language is fluent with minor grammatical errors such as 'molecve' instead of 'molecule' and 'tremend vice effects' instead of 'tremendous positive effects'. The content meets the Expected Output by summarizing a groundbreaking scientific finding.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 0.9

Reason: The Actual Output is highly relevant to the Input prompt, providing a comprehensive travel guide for Venice, Italy. It aligns well with the Expected Output by highlighting top attractions, accommodations, and activities. The content is coherent with a logical flow and consistent ideas. Minor issues in fluency such as 'Comprehasters' and 'enjoaching' slightly detract from readability.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.6

Reason: The Actual Output addresses the Input prompt and includes relevant details, but it lacks the exact date format and target audience. It also has minor formatting issues in the date and activities.

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 topic and includes key points such as the brain's role in coordinating body functions, processing external stimuli, and embodying consciousness. However, it uses overly complex terminology like 'preeminent neural center' and 'encapsulating the intricate

constructs,' which may hinder readability and coherence compared to the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

Model: AI.Models.phi3_mini-128k_4b_Q4

File: text_gen_001_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input prompt and aligns with the Expected Output by extracting the key information correctly. It is coherent and fluent with appropriate grammar and syntax. However, the format slightly differs from the Expected Output, which uses JSON formatting.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_002_result.json

Score: 0.7

Reason: The Actual Output addresses the main points but uses 'backyard' instead of 'garden' and adds unnecessary details. It is coherent and fluent but deviates slightly from the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_003_result.json

Score: 0.6

Reason: The Actual Output is relevant to the Input prompt and explains photosynthesis in a way that is engaging for a younger audience. However, it is overly complex and lengthy compared to the Expected Output, which is simpler and more direct. The Actual Output also includes some confusing analogies that may not be easily understood by children aged 8-10.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 0.5

Reason: The Actual Output is relevant to the Input and covers key features like noise-cancelling, 20-hour battery life, and comfort. However, it is overly verbose and repetitive, lacking the concise and professional tone seen in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.9

Reason: The Actual Output directly addresses the topic of a space adventure and aligns well with the Expected Output of astronauts exploring a new planet. The story is coherent and flows logically, with

detailed descriptions of the journey and discoveries. It is also fluent with correct grammar and natural language usage. However, it slightly deviates by including additional elements like an impending galactic disaster.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_006_result.json

Score: 0.8

Reason: The Actual Output addresses the topic of nature's beauty and covers key points with vivid imagery and personification. However, it includes commentary and suggestions for improvement, which disrupts the coherence and flow.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the topic of a recent scientific discovery, specifically a potential cure for Alzheimer's disease. It covers all key points and details expected in an informative article, including the discovery, key researchers, and the scientific basis of the findings. The article is coherent with a logical flow and consistent ideas, and it is fluent with correct grammar and natural language use.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 0.9

Reason: The Actual Output directly addresses the Input prompt by providing a comprehensive travel guide for Paris, which matches the context of the Expected Output. It is coherent, logically flowing from one section to the next, and maintains a consistent theme. The language is fluent, grammatically correct, and readable. It meets the anticipated quality and content standards by highlighting top attractions, accommodations, and activities. However, a minor typo ('thy' instead of 'the') slightly affects the overall perfection.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.5

Reason: The Actual Output addresses the topic and extracts relevant information but includes unnecessary details and lacks coherence. It misses the concise format of the Expected Output and adds speculative content not present in the Input.

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 relevant to the Input prompt, addressing the complexity and functions of the brain in a technical manner. It is coherent and fluent, with logical connections and correct grammar. The key elements and intent of the prompt are met, though some additional technical details were added beyond the Expected Output.

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 addresses the Input prompt and covers the key points mentioned in the Expected Output. However, it does not follow the JSON format required in 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 addresses the topic and is coherent and fluent. However, it misses the keyword 'playful' and uses 'playing' instead of 'exploring'.

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 topic of photosynthesis and adapts it for a younger audience. It covers the key points of plants using sunlight to create energy and relates it to how kids use food for energy. The explanation is coherent, engaging, and grammatically correct. However, it includes additional information about fruits and flowers not present in the Expected Output.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_004_result.json

Score: 1.0

Reason: The Actual Output directly addresses the topic by describing the Wireless Bluetooth Headphones and their features. It covers all key points such as noise-cancelling, 20-hour battery life, and comfortable fit, aligning well with the Expected Output. The ideas flow logically, and the text is coherent, grammatically correct, and readable.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_005_result.json

Score: 0.7

Reason: The Actual Output addresses the Input prompt by providing a creative space adventure story, staying on topic. However, it deviates from the Expected Output, which specifies a story about astronauts exploring a new planet. The coherence and fluency of the Actual Output are strong, with a logical flow and correct grammar.

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 topic of the beauty of nature, covering all key points expected in the poem. It is coherent, with a logical flow and consistent ideas, and is fluent with correct grammar and natural language use.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_007_result.json

Score: 1.0

Reason: The Actual Output directly addresses the topic of a recent scientific discovery by detailing a breakthrough in bacterial genome sequencing. It covers all key points such as the techniques used, the implications for antibiotic development, and quotes from researchers, matching the Expected Output's requirement for an informative article. The article is coherent, with a logical flow and clear connections between ideas, and it is fluent, free of grammatical errors.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_008_result.json

Score: 1.0

Reason: The Actual Output directly addresses the topic of creating a travel guide for a popular tourist destination, specifically Tokyo. It covers key points such as top attractions, local transportation, food experiences, nightlife, and practical tips, aligning well with the Expected Output. The content is coherent, logically structured, and fluently written with proper grammar and syntax.

Prompt_language: en

Expected_answer_language: en

Generated_text_language: en

File: text_gen_009_result.json

Score: 0.9

Reason: The Actual Output addresses the Input prompt and includes all key details such as event name, dates, and location. It also mentions event components and target audience, aligning well with the Expected Output. The output is coherent, well-organized, and fluent, with minor differences in format.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en

File: text_gen_010_result.json

Score: 0.9

Reason: The Actual Output addresses the Input prompt and aligns well with the Expected Output, providing a technical explanation of the brain's functions. It is coherent, fluent, and maintains a logical flow. However, it is more detailed and expansive than the Expected Output, which is more concise.

Prompt_language: en
Expected_answer_language: en
Generated_text_language: en