



SpiritRAG: A Q&A System for Religion and Spirituality in the United Nations Archive

Yingqiang Gao^{†UZH^{LIRI}} Fabian Winiger^{‡UZH^{TRF}}

Patrick Montjourides^{‡UZH^{IFE}} Anastassia Shaitarova^{‡UZH^{LIRI}}

Nianlong Gu^{UZH^{LIRI}} Simon Peng-Keller^{UZH^{TRF}} Gerold Schneider^{UZH^{LIRI}}

^{UZH^{LIRI}} Linguistic Research Infrastructure, University of Zurich, Switzerland

^{UZH^{TRF}} Faculty of Theology and the Study of Religion, University of Zurich, Switzerland

^{UZH^{IFE}} Institute of Education, University of Zurich, Switzerland

{yingqiang.gao, fabian.winiger}@uzh.ch

<https://SpiritRAG.linguistik.uzh.ch>

Abstract

Religion and spirituality (R/S) are complex and highly domain-dependent concepts which have long confounded researchers and policymakers. Due to their context-specificity, R/S are difficult to operationalize in conventional archival search strategies, particularly when datasets are very large, poorly accessible, and marked by information noise. As a result, considerable time investments and specialist knowledge is often needed to extract actionable insights related to R/S from general archival sources, increasing reliance on published literature and manual desk reviews. To address this challenge, we present SpiritRAG, an interactive Question Answering (Q&A) system based on Retrieval-Augmented Generation (RAG). Built using 7,500 United Nations (UN) resolution documents related to R/S in the domains of health and education, SpiritRAG allows researchers and policymakers to conduct complex, context-sensitive database searches of very large datasets using an easily accessible, chat-based web interface. SpiritRAG is lightweight to deploy and leverages both UN documents and user provided documents as source material. A pilot test and evaluation with domain experts on 100 manually composed questions demonstrates the practical value and usefulness of SpiritRAG.*



Dataset



Code

1 Introduction

The concepts of “religion” and “spirituality” (R/S) are complex and highly domain-dependent, and vary substantially across social and cultural contexts. The meaning and connotation of R/S poses

*Corresponding author.

‡Equal contribution.

*SpiritRAG is under MIT license.

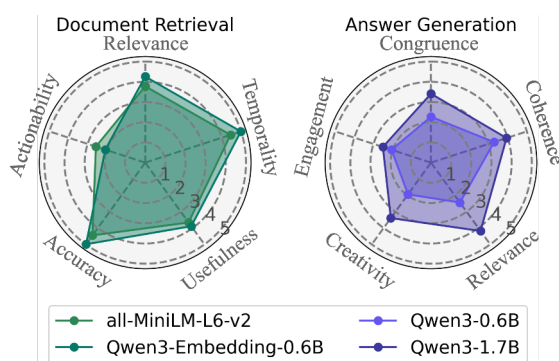


Figure 1: Human evaluation results of SpiritRAG on 50 health + R/S test questions listed in Appendix A. Our best-performing system setup received promising assessments from domain experts.

unique challenges to historians, sociologists, anthropologists, and other scholars who seek to understand R/S across distinct social and historical settings. Definitions of R/S vary widely across academic disciplines, and scholars have yet to reach a consensus on how to define them in general terms (Steenland et al., 2022; Winiger et al., 2025).

For policymakers, the complexity and domain-dependence of these terms are significant barriers to research and decision-making. This is illustrated by the long and protracted history of R/S in the United Nations (UN) system, the main locus of intergovernmental cooperation in the pursuit of global values for humanity, where R/S regularly surfaces as a matter of mutual concern in the areas of healthcare (Peng-Keller et al., 2022), Mental Health and Psychosocial Support (MHPSS) in humanitarian settings (Ager et al., 2019), soft diplomacy and value politics (Stensvold, 2017; Carrette, 2017; Baumgart-Ochse and Wolf, 2019; Steiner and Christie, 2021), and various develop-

ment and peace building efforts (UN Inter-Agency Task Force on Religion and Development, 2020). Accessing and analyzing documents produced by the organizations and agencies of the UN often require extended and time-consuming archival research using poorly designed or outdated search interfaces. This is exemplified by the official archive of the UN², which includes a vast collection of documents produced by the Security Council, the General Assembly, the Economic and Social Council (ECOSOC) and their subsidiaries, and others. Additionally, some UN entities, such as the World Health Organization (WHO) and the United Nations Educational, Scientific, and Cultural Organization (UNESCO), maintain separate archives.

The problem of accessibility is greatly exacerbated when searching for information on relatively niche, complex and domain-dependent concepts such as R/S, in which case the retrieval of actionable information requires a substantial investment of time to manually search very large numbers of often largely irrelevant documents marked by information “noise” such as procedural formalities, preambular clauses, extensive appendices, and frequently used phrases which contain keywords used in an unrelated context (e.g., “in the spirit of [...]”). The combination of large, poorly accessible and noisy databases with search queries related to complex and domain-dependent topics such as R/S underscores the urgent need for systems that can quickly and easily retrieve, synthesize and generate knowledge based on complex and nuanced user queries, providing scholars and policymakers with low-barrier access to accurate and reliable information on complex topics.

Current Large Language Models (LLMs) and general-purpose Information Retrieval (IR) systems often produce outputs that lack contextual sensitivity, thereby risking the homogenization of nuanced perspectives exemplified by the discourse on R/S (Hutchinson, 2024; Liu et al., 2024; Plaza-del Arco et al., 2024). Recent studies have highlighted that LLMs may exhibit contextual biases, particularly when predicting public opinion across diverse national and linguistic settings, leading to inaccuracies and a failure to capture cultural subtleties (Qu and Wang, 2024; von der Heyde et al., 2024). Furthermore, evaluations of LLMs’ cultural alignment reveal significant limitations, including instability across different cultural dimensions and

challenges in consistently steering models to represent specific cultural perspectives (Masoud et al., 2025; Khan et al., 2025). These findings underscore the need for specialized knowledge systems capable of retrieving and generating knowledge in information environments marked by high context- and domain dependence and cultural complexity. However, no existing systems were designed for knowledge acquisition in the context of R/S.

To address this gap, we present SpiritRAG—a Retrieval-Augmented Generation (RAG; Lewis et al. (2020)) Question Answering (Q&A) system designed to retrieve, synthesize, and generate R/S-related knowledge based on the resolutions stored in the UN archive. For present purposes, the corpus was built using data on health and education, with the latter included because many education-related documents are implicitly related to health (e.g., on tobacco control or sexual and reproductive care, see Sustainable Development Goal 3 (United Nations, 2024)). SpiritRAG is built on the top of approx. 7,500 documents available in seven languages, and supports interactive querying with real-time answers linked to exact reference documents. To evaluate the system, we compiled an expert-written test set of R/S-related questions in health and education. Pilot tests show that SpiritRAG is perceived by social scientists as a highly innovative and potentially groundbreaking interactive knowledge engine, highlighting its potential as a proof-of-concept for future LLM-augmented work in public policy, historical scholarship, and more generally, in the field of digital humanities.

Our main contributions are: (1) We developed an interactive Q&A system for retrieving R/S-related knowledge; (2) We built a corpus of nearly 7,500 UN resolution documents on health and education subjects; (3) We created a test set of 100 evaluation questions authored by domain experts.

A video demonstration of SpiritRAG can be accessed at https://youtu.be/x_G8bopwp8A?si=irrrlslw_i_ryRCX.

2 SpiritRAG: A Q&A System for R/S

SpiritRAG adopts a similar modular design principle as proposed by Gao et al. (2024) (see Figure 2). In the following, we discuss the core components of the proposed system.

2.1 Corpus

We constructed the UN-RES corpus by crawling the United Nations Official Document System² us-

²<https://documents.un.org/>

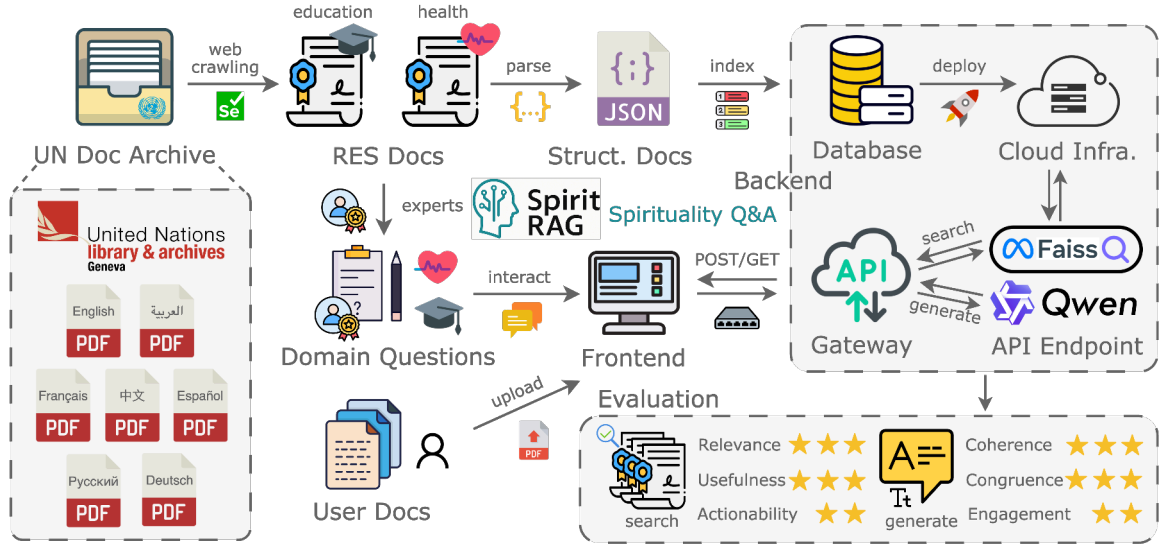


Figure 2: Architecture of SpiritRAG. SpiritRAG is an interactive Question Answering (Q&A) system enhanced by Retrieval-Augmented Generation (RAG) and specifically designed for acquiring domain-dependent knowledge in the resolution documents (RES Docs) of United Nations (UN) with a focus on Religion and Spirituality (R/S) related to health and education.

ing a Python script with the web browsing simulator Selenium³. Specifically, we searched for UN documents (PDF files) that are: (1) resolution documents (marked RES); (2) related to the subjects *health* and *education*; and (3) published between *01.01.1990* and *03.31.2025*. This process resulted in a total of nearly 7,500 UN resolution documents across seven languages: *Arabic* (ar), *Chinese* (zh), *English* (en), *French* (fr), *German* (de), *Russian* (ru), and *Spanish* (es). The overall dataset statistics are presented in Table 1. Figure 3 visualizes the temporal distribution of UN resolution documents within the health and education domains, along with the corresponding counts of unique subjects.

Freq.	ar	de	en	es	fr	ru	zh
health + R/S							
# Doc.	4,669	2,328	4,781	4,670	4,747	4,690	4,759
# Subj.	2,360	2,046	2,383	2,361	2,377	2,367	2,380
education							
# Doc.	2,586	1,123	2,718	2,658	2,694	2,670	2,586
# Subj.	1,865	1,455	1,944	1,919	1,935	1,930	1,938

Table 1: Frequency of documents and subjects per language for each subject domain.

To provide an initial overview of the data, we conducted a hierarchical cluster analysis of the 2,652 unique subject tags included in the metadata of the UN-RES corpus. The resulting 185 concise and interpretable clusters allowed for deeper in-

³<https://github.com/seleniumhq/selenium>, Apache-2.0 license.

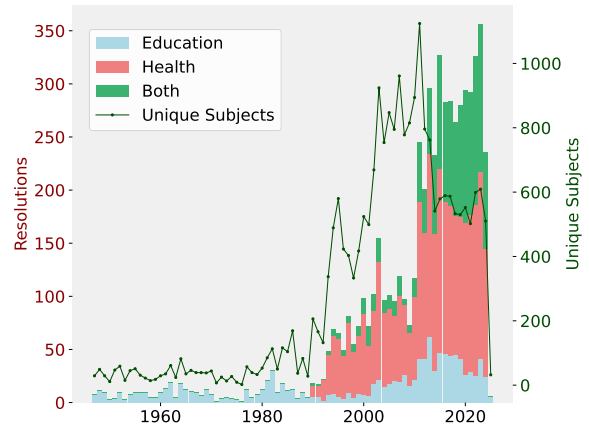


Figure 3: Resolutions and subjects across domains and decades in the UN-RES dataset.

sights into the temporal distribution of religious themes within the contexts of health and education. Figure 7 in Appendix D shows the average weight fluctuations of selected relevant clusters across decades. Figure 8 presents a more fine-grained analysis of specific religion-related subjects, broken down by five-year periods.

We further parsed crawled documents in English, French, German, and Spanish into a structured JSON format with paragraph chunks using the document parsing toolkit Docling⁴ (Deep Search Team, 2024)⁵. From the parsed documents, we con-

⁴<https://github.com/docling-project/docling>, MIT license.

⁵We did not parse other languages due to the lack of high-

structured the vector database for UN-RES with the efficient nearest neighbor search toolkit `Faiss`⁶, by using the pre-trained text encoder SentenceBERT (Reimers and Gurevych, 2019) and Qwen3-Embedding (Zhang et al., 2025) to pre-compute the document embeddings.

2.2 Models

Document retrieval. Given a user query q (i.e., a R/S-related question), we use the same text encoder to compute the query embedding \mathbf{e}_q on the fly. The document retriever first computes the cosine similarity score between the query embedding \mathbf{e}_q to each pre-computed document embedding $\mathbf{e}_d \in \mathbf{E}^{|D| \times m}$ (where m denotes the embedding dimensionality, D the corpus, and d the document in D) stored in the database:

$$\text{cos_sim}(\mathbf{e}_q, \mathbf{e}_d) = \frac{\mathbf{e}_q \cdot \mathbf{e}_d}{\|\mathbf{e}_q\| \cdot \|\mathbf{e}_d\|}.$$

Based on the resulting similarity scores, the retriever pre-fetches the top- n most relevant resolution documents $(D_i)_{i=1}^n$ by ranking them in descending order of cosine similarity (i.e., the higher the similarity, the higher the rank):

$$\text{rank}(d) = |\{d' \in D \mid \text{cos_sim}(\mathbf{e}_q, \mathbf{e}_{d'}) > \text{cos_sim}(\mathbf{e}_q, \mathbf{e}_d)\}| + 1.$$

The document retriever then re-ranks the documents in $(D_i)_{i=1}^n$ by calculating the similarity between the query embedding and each sentence embedding \mathbf{e}_s within a given D_i , using both the maximal and average similarities. Each D_i is re-ranked according to the relevance score

$$r(D_i; q) = \alpha \cdot \max_{s \in D_i} (L_2(\mathbf{e}_q, \mathbf{e}_s)) + (1 - \alpha) \cdot \text{avg}_{s \in D_i} (L_2(\mathbf{e}_q, \mathbf{e}_s)),$$

where s is a sentence in D_i , L_2 denotes the Euclidean distance and α is a weighting factor that was empirically set to 0.7. The final retrieved documents are the top- k entries in the re-ranked list.

Answer generation. Conditioned on the user query and the top- k most relevant documents retrieved from the database, we use instruction-tuned Qwen3 models (Yang et al., 2025) to generate the answer by assembling the inputs with a simple prompt template (see Figure 4) and passing them to the model.

quality PDF parsers.

⁶<https://github.com/facebookresearch/faiss>, MIT license.

You are a helpful AI assistant. Use the following information to answer the user’s question.

User’s question: {query}

Relevant information from the retrieved documents: {retrieved_docs}

Relevant information from the user uploaded PDF (optional): {parsed_pdf}

Figure 4: Prompt template used for answer generation.

In practice, users often have reliable source documents that may provide additional context for their questions. To support this, we implemented a function that allows users to upload a source PDF file, which is parsed and chunked in real time. When available, this source file is also included in the conditional generation of the answer.

To accelerate the generation process, we implemented the answer generation with the efficient inference toolkit `vLLM`⁷ (Kwon et al., 2023).

2.3 Demonstration

SpiritRAG has a front-end implemented with `React.js`⁸ and a back-end engineered using `Flask`⁹ APIs. With this modular design, SpiritRAG is lightweight and easily deployed both locally or on cloud services, requiring minimal computation resources. Our pilot system was deployed on a university-internal cloud service using an NVIDIA Tesla L4 GPU card with 24 GB of RAM.

Figure 5 demonstrates the user interface of SpiritRAG. SpiritRAG allows real-time human-machine interactions in a chatbot setting, where users can freely ask questions related to R/S. The answers generated by the Qwen3 model are rendered in Markdown for improved readability. Users can switch to dark mode and download the conversation history by clicking on the buttons in the banner.

On the right panel of the web application, accessible via the source button in the lower right corner, the most relevant documents used for answer generation are displayed (see Figure 9a in

⁷<https://github.com/vllm-project/vllm>, Apache-2.0 license.

⁸<https://react.dev/>

⁹<https://flask.palletsprojects.com/>

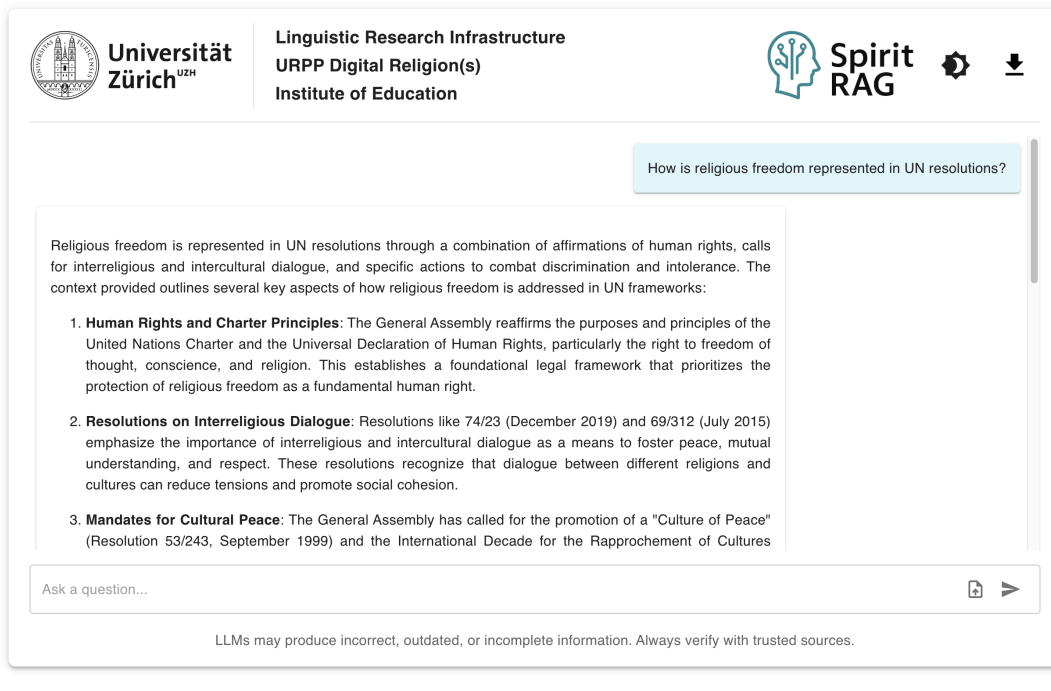


Figure 5: Overview of SpiritRAG. Generated responses are rendered in Markdown to improve readability of long answers. This example answer was generated by Qwen3-1.7B with Qwen3-Embedding-0.6B as document retriever.

Appendix E). Each retrieved document is presented with its metadata, including document title, key subjects (highlighted in green), publication date, and available languages. We report more statistics on the subjects in Appendix D.

When one of the available languages is clicked, a PDF viewer will be opened in the left panel, displaying the original UN resolution document (see Figure 9b in Appendix E). The left panel can be closed by clicking on any empty area on the web application.

3 Evaluation

We performed human evaluation of SpiritRAG’s document retrieval and answer generation capabilities. For this purpose, we asked two domain experts specialized in R/S and education research to assess SpiritRAG along multiple evaluation dimensions.

For each question in the test of 100 questions (see Appendix A), the domain experts were first asked to evaluate the retrieved documents. After an answer was generated, experts could activate the assessment mode by clicking the Eval button located at the top right of the right panel. For each retrieved document and the generated answer, a five-point Likert scale was provided to assess six evaluation dimensions (see Appendix A for details).

4 Results and Discussion

In Table 2 and Table 3, we report the average ratings across each evaluation dimension. We observe that the quality of both retrieved documents and generated answers improves as the backbone document retriever and answer generator are replaced with larger language models. Evaluators noted the considerable improvement in performance between Sentence-BERT and Qwen3-Embedding as document retriever, and particularly a significant and consistent improvement in answer length and detail of Qwen3-0.6B in comparison with Qwen3-1.7B as answer generator. It is worth noting that, in the case of document retrieval, a low actionability score does not necessarily indicate poor retriever performance. Rather, it reflects the fact that only a limited number of UN resolutions contain actionable directives, and ambivalence regarding how actionability should be understood in the context of several evaluation questions.

A more detailed, iterative strategy of evaluation and fine-tuning based on systematically developed evaluation dimensions tailored to the corpus was suggested to further strengthen the feasibility of SpiritRAG as a widely used research tool. An informal pilot in an interdisciplinary academic set-

Document Retriever	Relevance	Accuracy	Usefulness	Temporality	Actionability
all-MiniLM-L6-v2	3.78	4.45	3.66	4.45	2.57
Qwen3-Embedding-0.6B	4.26	4.99	3.89	4.99	2.08
Answer Generator	Congruence	Coherence	Relevance	Creativity	Engagement
Qwen3-0.6B	2.27	3.29	2.44	1.94	2.06
Qwen3-1.7B	3.41	3.94	4.18	3.39	2.51

Table 2: Human evaluation results on health + R/S related test questions.

Document Retriever	Relevance	Accuracy	Usefulness	Temporality	Actionability
all-MiniLM-L6-v2	2.56	2.57	2.44	4.70	2.38
Qwen3-Embedding-0.6B	3.51	3.43	3.42	4.75	3.39
Answer Generator	Congruence	Coherence	Relevance	Creativity	Engagement
Qwen3-0.6B	2.42	2.71	2.23	3.54	2.71
Qwen3-1.7B	3.78	3.76	3.98	3.92	4.06

Table 3: Human evaluation results on education related test questions.

ting suggested that SpiritRAG is perceived by subject experts as a highly innovative, interdisciplinary proof-of-concept with the potential for far-reaching impact in their respective academic disciplines.

5 Related Works

SpiritRAG is relevant to scholars in religious studies, health policy, and international relations. It is of particular interest to researchers interested in methodological innovation in the study of religion (Wright, 2022), the use of digital humanities methods in the study of religion (Hutchings, 2015) and the use of neural networks for text analysis (Kim et al., 2020; Suissa et al., 2022). SpiritRAG also provides a workable proof-of-concept for theologians interested in exploring new forms of dialogue-based interaction with religious texts (Kurlberg, 2020; Garner, 2021; Sutinen and Cooper, 2021). SpiritRAG complements computational analyses of large text corpora in fields that have mostly focused on topic modeling to date (Park et al., 2023; Yamada, 2024) and proposes a new resource for scholars looking at inter-textual discursive relationships in global governance and social policy (Windzio and Heiberger, 2022; Montjourides, 2023).

6 Conclusion and Future Work

We presented SpiritRAG, a Q&A system enhanced with RAG for the domains of health plus R/S within the context of United Nations resolution documents. Built on top of approx. 7,500 UN resolution documents, SpiritRAG is designed to support social scientists and policymakers in the initial stages of knowledge acquisition by enabling interactive retrieval of contextualized, domain-dependent infor-

mation. This, in turn, eases evidence-informed research and decision-making. In future work, we plan to expand SpiritRAG to incorporate a broader range of document sources, including religious texts, with the goal of making the embedded spiritual knowledge in these documents more accessible and relevant to contemporary discourse.

Limitations

SpiritRAG has several limitations. First, due to computational constraints, we did not evaluate larger-scale LLMs for document retrieval and answer generation. Second, in the absence of ground-truth answers and relevant documents, we did not fine-tune the document retriever for optimal performance. Third, SpiritRAG is currently tailored to address education- or health + R/S-related questions within the context of UN resolution documents only; it does not incorporate documentation from other UN organizations and related multilateral organizations. Nevertheless, thanks to its modular architecture, SpiritRAG can be readily extended to other domain-specific knowledge sources.

Acknowledgments

This work is funded by the URPP Digital Religion(s) at the University of Zurich (UZH) and the Swiss National Science Foundation, grant number 204286, with extensive support provided by the Chair of History of Education and Education Policy Analysis at Institute of Education of UZH. We also acknowledge NCCR Evolving Language, Swiss National Science Foundation Agreement No. 51NF40_180888. A special thanks to Mr. Igor Mustač at LiRI for deploying SpiritRAG online.

References

- Wendy Ager, Michael French, Atallah Fitzgibbon, and Alastair Ager. 2019. The Case for—and Challenges of—faith-sensitive Psychosocial Programming. *Intervention*, 17(1):69–75.
- Claudia Baumgart-Ochse and Klaus Dieter Wolf, editors. 2019. *Religious NGOs at the United Nations: Polarizers or Mediators?* Routledge studies in religion and politics. Routledge, Taylor & Francis Group, London.
- Jeremy R. Carrette. 2017. *Religion, NGOs and the United Nations: Visible and Invisible Actors in Power*. Religious studies. Bloomsbury Academic, London.
- Deep Search Team. 2024. Docling Technical Report.
- Yingqiang Gao, Jhony Prada, Nianlong Gu, Jessica Lam, and Richard HR Hahnloser. 2024. MODOC: A Modular Interface for Flexible Interlinking of Text Retrieval and Text Generation Functions. *arXiv preprint arXiv:2408.14623*.
- Stephen Garner. 2021. Theology and the New Media. In *Digital religion*, pages 266–281. Routledge.
- Tim Hutchings. 2015. Digital Humanities and the Study of Religion. In Patrik Svensson and David Theo Goldberg, editors, *Between Humanities and the Digital*, pages 283–294. The MIT Press, Massachusetts.
- Ben Hutchinson. 2024. Modeling the Sacred: Considerations when Using Religious Texts in Natural Language Processing. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 1029–1043.
- Ariba Khan, Stephen Casper, and Dylan Hadfield-Menell. 2025. Randomness, not Representation: The Unreliability of Evaluating Cultural Alignment in LLMs. *arXiv preprint arXiv:2503.08688*.
- Seong-Hyeon Kim, Narae Lee, and Pamela Ebstyn King. 2020. Dimensions of Religion and Spirituality: A Longitudinal Topic Modeling Approach. *Journal for the scientific study of religion*, 59(1):62–83.
- Jonas Kurlberg. 2020. Doing God in Digital Culture: How Digitality is Shaping Theology. *Cursor_ Zeitschrift für explorative Theologie*. <https://cursor.pubpub.org/pub/sk6clz8d>.
- Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient Memory Management for Large Language Model Serving with PagedAttention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, and 1 others. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in neural information processing systems*, 33:9459–9474.
- Songyuan Liu, Ziyang Zhang, Runze Yan, Wei Wu, Carl Yang, and Jiaying Lu. 2024. Measuring Spiritual Values and Bias of Large Language Models. *arXiv preprint arXiv:2410.11647*.
- Reem Masoud, Ziquan Liu, Martin Ferianc, Philip C Treleaven, and Miguel Rodrigues Rodrigues. 2025. Cultural Alignment in Large Language Models: An Explanatory Analysis Based on Hofstede’s Cultural Dimensions. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 8474–8503.
- Patrick Montjourides. 2023. *Is this the Future We Want? Understanding the Legitimacy of International Education Agendas. The Example of Equity in Education*. Ph.D. thesis, Apollo - University of Cambridge Repository. DOI: 10.17863/CAM.89624.
- Kyung Hee Park, Qi Liu, and He Li. 2023. UNESCO Practices under the SDG4 and COVID-19 Response Framework: Topic Modeling from 2003 to 2021. *International Journal of Innovative Research and Scientific Studies*, 6(2):322–329.
- Simon Peng-Keller, Fabian Winiger, and Raphael Rauch. 2022. *The Spirit of Global Health: The World Health Organization and the “Spiritual Dimension” of Health, 1946-2021*. Oxford University Press, Oxford, New York.
- Flor Plaza-del Arco, Amanda Curry, Susanna Paoli, Alba Cercas Curry, and Dirk Hovy. 2024. Divine LLaMAs: Bias, Stereotypes, Stigmatization, and Emotion Representation of Religion in Large Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 4346–4366.
- Yao Qu and Jue Wang. 2024. Performance and Biases of Large Language Models in Public Opinion Simulation. *Humanities and Social Sciences Communications*, 11(1):1–13.
- Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992.
- Brian Steensland, Jaime Kucinkas, and Anna Sun. 2022. *Eminently Social Spirituality*. Oxford University Press New York.
- Sherrie M. Steiner and James T. Christie, editors. 2021. *Religious Soft Diplomacy and the United Nations: Religious Engagement as Loyal Opposition*. Lexington Books, Lanham.

- Anne Stensvold. 2017. *Religion, State and the United Nations: Value Politics*. Routledge Studies in Religion and Politics. Taylor and Francis, London and New York.
- Omri Suissa, Avshalom Elmalech, and Maayan Zhitomirsky-Geffet. 2022. Text Analysis Using Deep Neural Networks in Digital Humanities and Information Science. *Journal of the Association for Information Science and Technology*, 73(2):268–287.
- Erkki Sutinen and Anthony-Paul Cooper. 2021. What is Digital Theology? In *Digital Theology: A Computer Science Perspective*, pages 13–32. Emerald Publishing Limited.
- UN Inter-Agency Task Force on Religion and Development. 2020. Engaging with Religion and Faith-based Actors on the 2030 Sustainable Development Agenda - 2020. Technical report.
- United Nations. 2024. The Sustainable Development Goals Report 2024.
- Leah von der Heyde, Anna-Carolina Haensch, and Alexander Wenz. 2024. United in Diversity? Contextual Biases in LLM-Based Predictions of the 2024 European Parliament Elections. *arXiv preprint arXiv:2409.09045*.
- Michael Windzio and Raphael Heiberger. 2022. *Talking About Education: How Topics Vary Between International Organizations*, pages 239–266. Springer International Publishing, Cham.
- Fabian Winiger, Gerold Schneider, Janis Goldzycher, David Neuhold, and Simon Peng-Keller. 2025. The ‘Spiritual’ and the ‘Religious’ in the Twittersphere: A Topic Model and Semantic Map. *Journal of Religion, Media and Digital Culture*, 14(1):1–22.
- Bradley R. E. Wright. 2022. Methodological Innovations for the Study of Spirituality. In *Situating Spirituality - Context, Practice, and Power*, pages 113–127. Oxford University Press.
- Shoko Yamada. 2024. The Synchronic and Diachronic Evolution of Key Themes around SDG 4 before and after 2015: From a Quantitative Analysis of Web-downloaded Texts. *International Review of Education*, 70(4):651–671.
- An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 Technical Report. *arXiv preprint arXiv:2505.09388*.
- Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, and 1 others. 2025. Qwen3 Embedding: Advancing Text Embedding and Reranking Through Foundation Models. *arXiv preprint arXiv:2506.05176*.

A Test Set Questions and Evaluation Dimensions

ID	Question	Type	TB?
1	How has the relationship between religion, spirituality, and health been framed in policy texts adopted since 1946?	Open-ended	Yes
2	What references are made to gender equality in relation to religion, spirituality across UN resolutions?	Open-ended	No
3	Which resolutions cite religion and/or spirituality?	List (finite)	No
4	Have any UN resolutions addressed the right to religious freedom?	Yes/No	No
5	What provisions have been made for refugees' religious freedom?	Open-ended*	No
6	How is the phrase "culturally appropriate" used across UN documents?	Open-ended*	No
7	Do UN documents address the notion of human dignity?	Yes/No	No
8	What emphasis is placed on the notion of human dignity?	Open-ended	No
9	When was United Nations Interagency Task Force on Religion and Sustainable Development incorporated in UN resolutions?	Closed	Yes
10	What UN Member States commitments have been made toward religious freedom?	Open-ended*	No
11	How often is religious freedom mentioned in UN resolutions since 1990?	Closed	No
12	How has UN language associated with religion and/or spirituality evolved over time?	Open-ended*	Yes
13	What international monitoring frameworks are proposed for religious freedom-related targets?	List (finite)	Yes
14	How is religious freedom represented in UN resolutions?	Open-ended	Yes
15	Have there been calls for interreligious or interfaith dialogue in UN texts and resolution?	Yes/No	No
16	Are there any specific resolutions that discuss religion or spirituality?	List (finite)	No
17	How does the UN address religion and/or spirituality in conflict-affected regions?	Open-ended	Yes
18	Which resolutions link religion and/or spirituality to UN activities?	List (finite)	No
19	Are there references to faith-based organizations in UN resolutions?	Yes/No	No
20	How have UN texts and resolutions evolved in cultural appropriateness?	Open-ended*	Yes
21	How have religious leaders, communities or actors contributed to the Sustainable Development Goals?	Open-ended*	No
22	Does any UN text or resolution cite cultural appropriateness as a priority?	Yes/No	No
23	Is there any reference to religious involvement in UN-related activities in recent UN texts and resolutions?	Yes/No	Yes
24	Is interfaith or interreligious dialogue addressed by the United Nations?	Yes/No	No
25	What resolutions discuss interfaith or interreligious education?	List (finite)	No
26	Have any resolutions referred to culturally appropriate policies?	Yes/No	No
27	Is health framed as a human right consistently?	Open-ended	No
28	What are the main dimensions of religious discrimination acknowledged in UN resolutions?	List (finite)	Yes
29	Does any UN organizations take a specific interest in religion and/or spirituality?	Yes/No	No
30	How well do UN texts and resolutions address issues pertaining to public expenditure on healthcare?	Open-ended*	No
31	How do UN resolution approach healthcare for persons in low- or middle income countries?	Open-ended*	No
32	What frameworks have been proposed over time to monitor equity in healthcare?	List (finite)	No

Table 4: Health + R/S test questions (part 1).

ID	Question	Type	TB?
33	How often has “universal primary healthcare” as a global or priority been emphasized in the past decade?	Open-ended	Yes
34	Is “Health for All” referenced in UN resolutions?	Yes/No	No
35	What kind of healthcare system reforms are promoted by United Nations organizations?	Open-ended*	No
36	What role is assigned to UN agencies like WHO in healthcare?	Open-ended*	No
37	How is health in emergencies addressed by the United Nations?	Open-ended	Yes
38	Do any resolutions define global priorities for healthcare? And if so, how are they defined?	Yes/No + Open-ended	No
39	What provisions exist for global health emergencies in UN official texts?	Open-ended	No
40	Are health disparities based on socioeconomic status addressed by the UN?	Yes/No	No
41	How have the United Nations addressed racial or ethnic discrimination in healthcare over time?	Open-ended*	No
42	Have the United Nations paid attention to healthcare in rural areas?	Yes/No	No
43	To which extent do UN resolutions support healthcare education?	Open-ended	No
44	How have the UN embedded healthcare in their mandate to foster peaceful societies?	Open-ended	No
45	What resolutions address data collection in healthcare systems?	List (finite)	No
46	Are there mentions of global pandemics in health?	Yes/No	No
47	How do resolutions define the right to health?	Open-ended	Yes
48	What role does healthcare play in sustainable development according to the UN?	Open-ended	Yes
49	Are gender-based discrimination in healthcare among the concerns voiced by UN Member States?	Yes/No	No
50	What resolutions mention health as a human right?	List (finite)	No

Table 5: Health + R/S test questions (part 2).

ID	Question	Type	TB?
1	How has the right to education been framed in UN texts adopted since 1946?	Open-ended	Yes
2	What references are made to gender equality in education across UN resolutions?	Open-ended	No
3	Which resolutions discuss access to primary education?	List (finite)	No
4	Have any UN resolutions addressed the digital divide in education?	Yes/No	No
5	What provisions have been made for refugees’ education?	Open-ended*	No
6	How is inclusive education defined in UN resolutions?	Open-ended*	No
7	Do UN resolutions address special needs education?	Yes/No	No
8	What emphasis is placed on technical and vocational education in UN documents between 1960 and 1970?	Open-ended	Yes
9	When was lifelong learning incorporated in UN resolutions?	Closed	No
10	What UN Member States commitments have been made toward teacher training and professional development?	Open-ended*	No
11	How often is early childhood education mentioned in UN resolutions since 1981?	Closed	Yes
12	How has the UN language associated with justice and fairness in education evolved over time?	Open-ended*	No
13	What international monitoring frameworks are proposed for education-related targets?	List (finite)	Yes
14	How are indigenous peoples’ education rights represented in UN resolutions?	Open-ended	Yes
15	Have there been calls for universal free education in UN texts and resolution?	Yes/No	No
16	Are there any specific resolutions that define quality education?	List (finite)	No
17	How does the UN address education in conflict-affected regions?	Open-ended	Yes

Table 6: Education-related test questions (part 1).

ID	Question	Type	TB?
18	Which resolutions link education and peace building?	List (finite)	No
19	Are there references to education financing in UN resolutions?	Yes/No	No
20	How have UN texts and resolutions evolved in support for girls' education?	Open-ended*	No
21	How has the SDG 4 agenda been further reflected in resolutions since 2015?	Open-ended*	Yes
22	Does any UN text or resolution cite learning outcomes as a priority?	Yes/No	No
23	Is there any reference to private sector involvement in education in recent UN texts and resolutions?	Yes/No	Yes
24	Are online and distance learning modalities addressed by the United Nations?	Yes/No	No
25	What resolutions discuss multilingual education?	List (finite)	No
26	Have any resolutions referred to climate change education?	Yes/No	No
27	Is education framed as a human right consistently?	Open-ended	No
28	What are the main dimensions of exclusion acknowledged in UN resolutions during the Cold War?	List (finite)	Yes
29	Does UNESCO recommend specific pupils-teacher ratios?	Yes/No	No
30	How well do UN texts and resolutions address issues pertaining to public expenditure on education?	Open-ended*	No
31	How do UN resolutions approach education for persons with disabilities?	Open-ended*	No
32	What are the frameworks proposed in UN resolution to monitor progress towards equity in education?	List (finite)	No
33	How often has literacy as a goal or priority been emphasized in the past decade?	Open-ended	Yes
34	Is citizenship or civic education referenced in UN resolutions?	Yes/No	No
35	What kind of education system reforms are promoted by the United Nations?	Open-ended*	No
36	What role is assigned to UN agencies like UNESCO in education?	Open-ended*	No
37	How is education in emergencies addressed by the United Nations?	Open-ended	Yes
38	Do any resolutions define global citizenship education? And if so, how is it defined?	Yes/No + Open-ended	No
39	What provisions exist for non-formal education in UN official texts?	Open-ended	No
40	Are educational disparities based on socioeconomic status addressed by the UN?	Yes/No	No
41	How have the United Nations addressed racial or ethnic discrimination in education over time?	Open-ended*	No
42	Have the United Nations paid attention to education in rural areas?	Yes/No	No
43	To which extent do UN resolutions support bilingual education?	Open-ended	No
44	How have the UN embedded education in their mandate to foster peaceful societies?	Open-ended	No
45	What resolutions address data collection in education systems?	List (finite)	No
46	Are there mentions of international assessments in education?	Yes/No	No
47	How do resolutions define access to education?	Open-ended	Yes
48	What role does education play in sustainable development according to the UN?	Open-ended	Yes
49	Are gender stereotypes in curricula among the concerns voiced by UN Member States?	Yes/No	No
50	What resolutions mention teacher rights and working conditions?	List (finite)	No

Table 7: Education-related test questions (part 2).

Type	Definition
List (finite)	Desired answers are in form of finite list with bullet points.
Yes/No	Questions that can be answered in binary form.
Closed	Questions that can be precisely answered.
Open-ended	No special format requirement for desired answers.
Open-ended*	No special format requirement for desired answers, yet diachronic changes possible.

Table 8: Definition of question types.

Dimension	Descriptions
Document Retrieval	
Relevance	How relevant is the retrieved document to the question?
Accuracy	How trustworthy is the retrieved document for answering the question?
Usefulness	How useful is the retrieved document for answering the question?
Temporality	Is the retrieved document related to the time period implied by the question?
Actionability	How many actionable insights does the retrieved document contain?
Answer Generation	
Congruence	How well does the answer align with the retrieved documents?
Coherence	How logically consistent is the answer?
Relevance	How relevant is the answer to the question?
Creativity	How original and creative is the answer?
Engagement	How engaging and interesting is the answer?

Table 9: Evaluation dimensions for the retrieved documents and generated answer.

B Dataset Construction Details

We implemented the web crawler using Selenium in combination with BeautifulSoup¹⁰. Boolean operators were employed to combine keyword sets for document retrieval: *educat**, *school**, and *learn** for education-related documents, and *health*, *faith*, *religi**, *spiritual**, and *belief* for documents related to health and religion/spirituality (R/S). Notably, some resolution documents are relevant to both domains; see Figure 6 for an illustrative example.

Welcome to the United Nations								
A/RES/70/1 Download								
General Arabic Chinese English French Russian Spanish German Reference Explorer								
Title	Transforming our world : the 2030 Agenda for Sustainable Development : resolution / adopted by the General Assembly							
Symbol	A/RES/70/1							
Session / Year	70							
Agenda Item(s)	15 116							
Distribution	GEN							
Area	UNDOC							
Publication Date	21/10/2015							
Subject(s)	SUSTAINABLE DEVELOPMENT, DEVELOPMENT STRATEGIES, PEACE, PARTNERSHIP, POVERTY MITIGATION, HUNGER, FOOD SECURITY, HEALTH, EDUCATION, GENDER EQUALITY, WATER MANAGEMENT, SUSTAINABLE ENERGY, EMPLOYMENT, INDUSTRIALIZATION, INCOME DISTRIBUTION, HUMAN SETTLEMENTS, SUSTAINABLE CONSUMPTION, CLIMATE CHANGE, MARINE RESOURCES, TERRESTRIAL ECOSYSTEMS, JUSTICE							
Languages	ARABIC	CHINESE	ENGLISH	FRENCH	RUSSIAN	SPANISH	GERMAN	
Job Number	N1529187	N1529188	N1529189	N1529190	N1529192	N1529193	N1529191	
Release Date	28/02/2019	22/10/2015	22/10/2015	28/02/2019	22/10/2015	15/12/2015	03/03/2020	

Figure 6: An example resolution document that is both health- and education-related. Screenshot taken from the website of UN Official Document System.

¹⁰<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>, MIT License.

C Examples of Generated Answers

Q: What references are made to gender equality in relation to religion, spirituality across UN resolutions?

Qwen3-Embedding-0.6B for document retrieval and Qwen3-1.7B for answer generation

A: The references to gender equality in relation to religion, spirituality, and the promotion of interreligious and intercultural dialogue are embedded within several UN resolutions, particularly those addressing cultural pluralism, human rights, and the prevention of violence and discrimination. Here is a detailed breakdown of the key references:

1. Promotion of Interreligious and Intercultural Dialogue

Resolution 75/26 (adopted on 2 December 2020) emphasizes the importance of interreligious and intercultural dialogue for peace, mutual understanding, and social stability. It underscores the need to create an environment conducive to peace and mutual respect at both global and local levels. This resolution directly ties gender equality to religious and spiritual diversity, highlighting the role of dialogue in fostering inclusive societies.

2. Human Rights and Religious Minorities

Resolution 74/306 (adopted on 11 September 2020) recognizes the need to counter discrimination, hate speech, and xenophobia related to the COVID-19 pandemic. It reaffirms the primary responsibility of states to protect human rights, including the rights of religious minorities. This resolution explicitly links gender equality to the protection of religious minorities, ensuring that women and other marginalized groups are not disproportionately affected by discrimination based on religion or belief.

3. Declaration on the Elimination of All Forms of Intolerance and Discrimination Based on Religion or Belief

Resolution 36/55 (adopted on 25 November 1981) proclaims the **Declaration on the Elimination of All Forms of Intolerance and Discrimination Based on Religion or Belief**, which includes provisions for gender equality. The declaration emphasizes the importance of protecting the rights of individuals, including women, and promotes the eradication of discrimination based on religion or belief.

4. International Day of Living Together in Peace

Resolution 72/130 (adopted on 8 December 2017) designates 28 December as the International Day of Living Together in Peace. While not explicitly mentioning gender equality, this resolution underscores the importance of mutual respect and understanding, which are critical for gender equality in religious and spiritual contexts.

5. United Nations Alliance of Civilizations

Resolution 69/312 (adopted on 6 July 2015) reaffirms the **United Nations Alliance of Civilizations**, which promotes dialogue among civilizations, cultures, religions, and beliefs. This initiative aligns with the goal of fostering gender equality by encouraging respectful interactions between diverse groups, including religious and spiritual communities.

6. Strategic Initiatives and Campaigns

UNESCO's **Unite for Heritage Campaign** (launched in 2015) and the **Plan of Action to Safeguard Religious Sites** (prepared by the United Nations Alliance of Civilizations) highlight the importance of preserving cultural heritage and promoting tolerance. These initiatives indirectly support gender equality by ensuring that religious and spiritual practices are respected and accessible to all, including women and girls.

7. Combating Violent Extremism and Hate Speech

Resolution 73/285 (adopted on 2 April 2019) calls for the elimination of violence based on religion or belief and emphasizes the role of interreligious and intercultural dialogue in countering hate speech. This resolution supports gender equality by addressing systemic issues that contribute to discrimination and violence, particularly against women and girls.

8. International Day of Commemoration of the Victims of Acts of Violence Based on Religion or Belief

Resolution 73/296 (adopted on 28 May 2019) designates 22 August as the International Day of Commemoration of the Victims of Acts of Violence Based on Religion or Belief. This day highlights the need to address violence rooted in religious or spiritual discrimination, which disproportionately affects women and girls.

9. Global Migration and Cultural Diversity

Resolution 74/23 (adopted on 12 December 2019) acknowledges the growing importance of interreligious and intercultural dialogue in the context of migration. This resolution reinforces the idea that cultural diversity and the pursuit of cultural development are sources of mutual enrichment, which is essential for gender equality in a multicultural society.

10. Role of Freedom of Expression in Combating Intolerance

Resolution 73/296 also reiterates the importance of the **right to freedom of opinion and expression** in strengthening democracy and combating religious intolerance. This principle is central to gender equality, as it ensures that women and other marginalized groups can express their views and participate in societal discourse without fear of discrimination.

Conclusion:

The UN resolutions collectively emphasize the intersection of religion, spirituality, and gender equality through the promotion of interfaith dialogue, the protection of human rights, and the prevention of discrimination. They recognize that religious and spiritual diversity must be respected and integrated into policies that advance gender equality, social justice, and peace. By fostering mutual understanding and respect, these resolutions aim to create inclusive societies where all individuals, regardless of their religious or spiritual beliefs, can thrive.

all-MiniLM-L6-v2 for document retrieval and Qwen3-0.6B for answer generation

A: The references to gender equality and religion/spirituality in UN resolutions are found in the following context:

1. Mainstreaming a Gender Perspective

Resolution adopted by the Economic and Social Council on 7 June 2023 (Agenda item 12).

- It reaffirms its agreement from 1997/2 and recalls previous resolutions (e.g., 2011/6, 2012/24, 2013/16, etc.).
- It emphasizes the importance of mainstreaming a gender perspective into all policies and programs.

2. Gender Equality and the Beijing Declaration and Platform for Action

- The UN has reiterated that gender mainstreaming is a critical strategy for achieving gender equality and empowering women and girls.
- The outcomes of the twenty-third special session and the implementation of the Programme of Action of the International Conference on Population and Development are highlighted.

3. Global Commitments

- The UN recognizes the importance of gender equality and empowerment through various summits and conferences, including the 2005 World Summit, 2011 World Summit, and the 2015 development agenda.

These references collectively emphasize the role of gender equality and the integration of religious/spiritual aspects into UN policies and initiatives.

Table 10: Answers generated by SpiritRAG. The average response time is dependent on the system configuration. Our best system setup takes around 7 seconds to response.

D Statistics on Domain-related Subjects in UN Resolution Documents

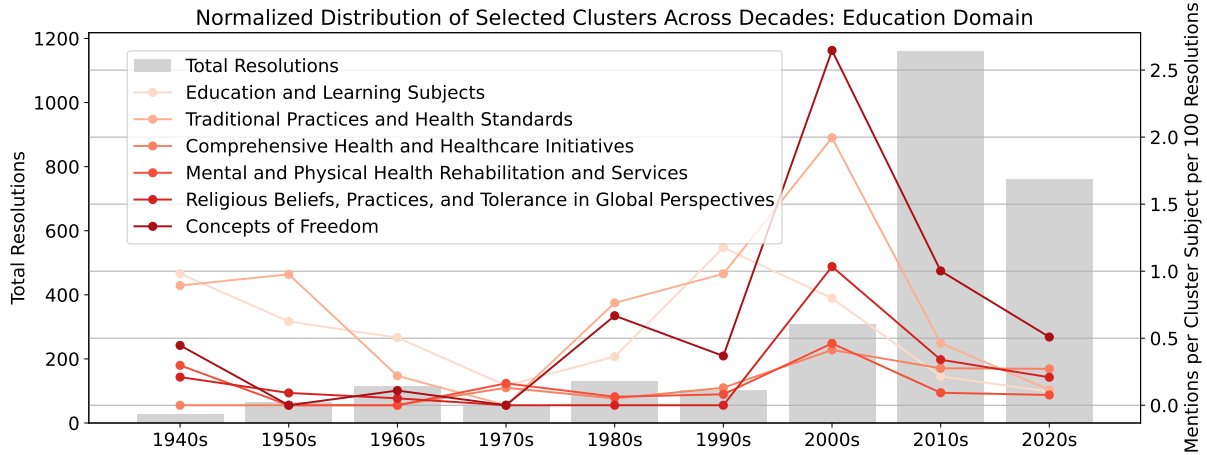


Figure 7: Temporal trends in the representation of selected subject clusters¹¹ within the education domain. The gray bars indicate the total number of education-related resolutions in each decade. Colored lines show the normalized frequency of mentions for each cluster, calculated as the average number of subject mentions per cluster (per 100 resolutions), adjusted for the number of subjects in each cluster.

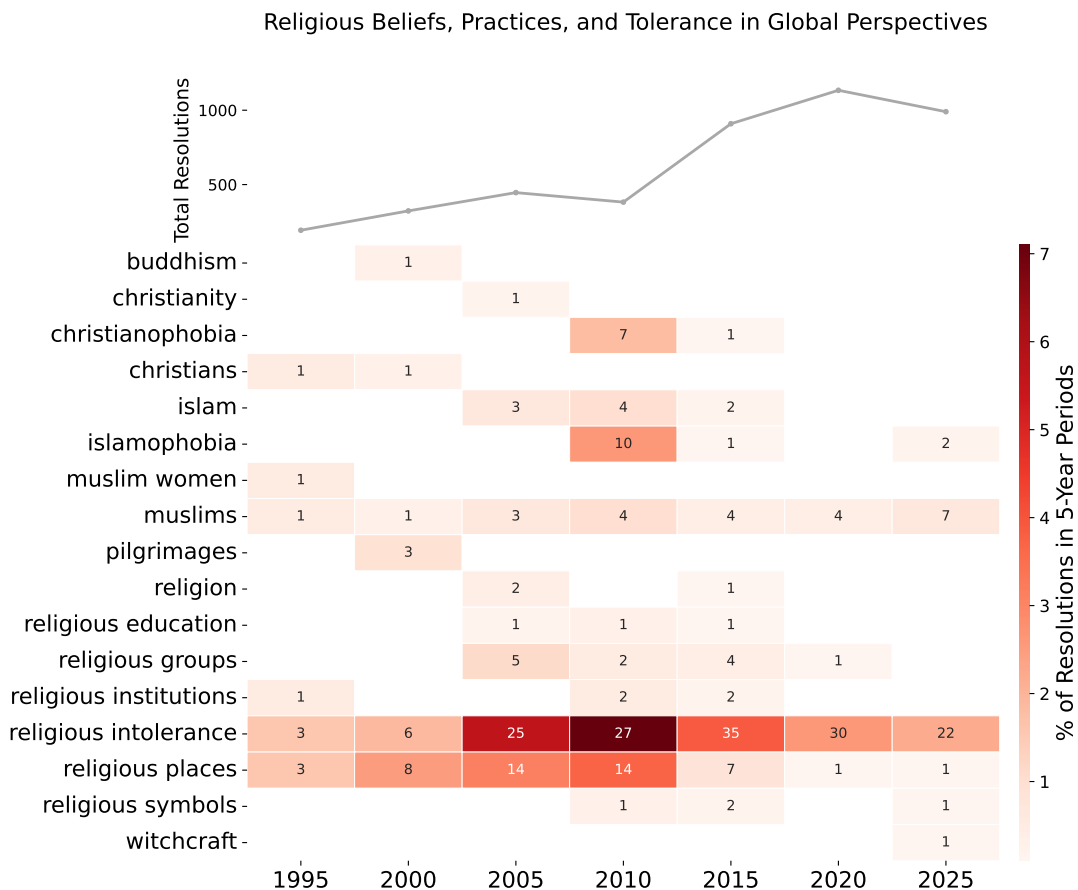
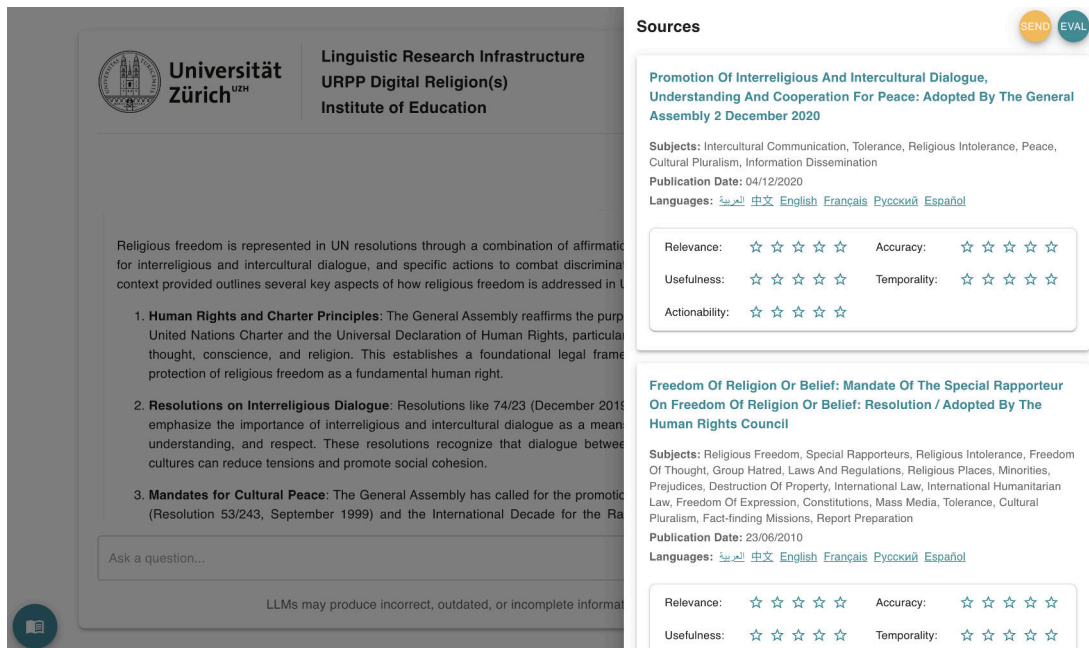


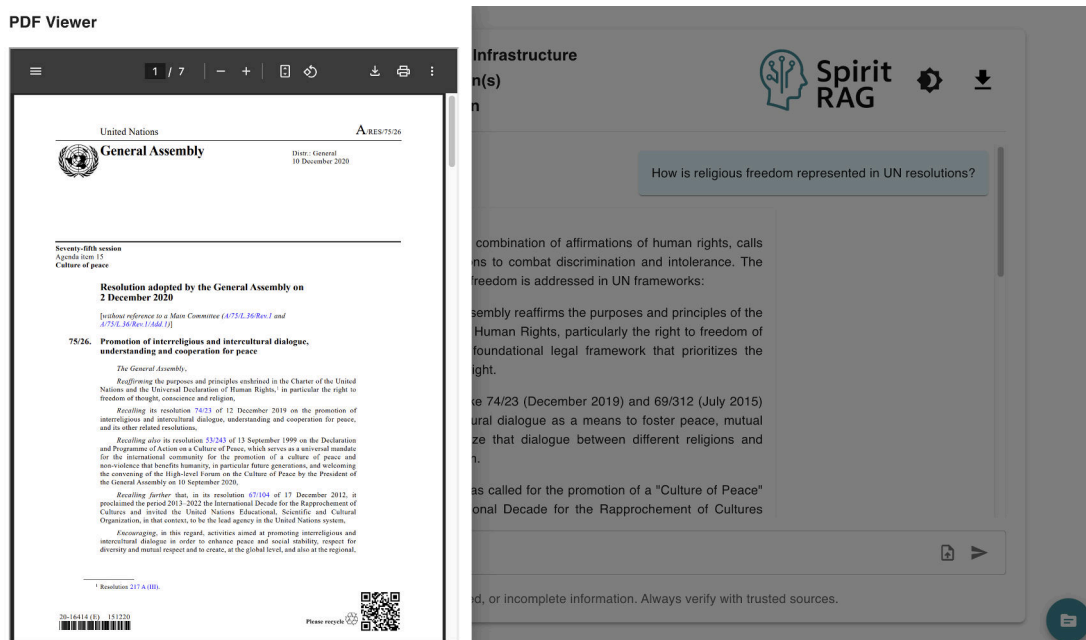
Figure 8: Distribution of the cluster #133 subjects within the health domain across 5-year periods. Cell color reflects the proportion of resolutions in each period tagged with a given subject. Raw counts of subjects are overlaid as annotations. The line chart above shows the total number of health-related resolutions in each period.

¹¹The clusters in Figures 7 and 8 were created using Sentence-BERT (all-MiniLM-L6-v2) and agglomerative clustering from sklearn with the distance threshold 2. Cluster titles were assigned automatically using GPT-4o.

E User Interface



(a) Retrieved documents.



(b) Document viewer.

Figure 9: User interface of SpiritRAG. Left panel (viewing original PDF document) and right panel (viewing the list of the most relevant documents to the user question).