

Adapting RAG to Specialised Communities: A UX-Driven Design Study

Eva Rolin*, Patrick Watrin*, Suzanne Kieffer*

* *Institute for Language and Communication (ILC) – Centre for Natural Language Processing (CENTAL)*

Université catholique de Louvain (UCLouvain)

Louvain-la-Neuve, Belgium

Emails: eva.rolin@uclouvain.be, patrick.watrin@uclouvain.be, suzanne.kieffer@uclouvain.be

Abstract—Large language models (LLMs) substantially transform information retrieval, with retrieval-augmented generation (RAG) emerging as a predominant paradigm. Although RAG demonstrates strong performance for generic information needs, its adaptation to specialized user communities remains insufficiently understood. In this paper, we address this issue through a collaboration with the *Académie royale de Belgique*, focusing on the design of a RAG-based search engine for the *Biographie nationale de Belgique*, a scholarly biographical corpus that serves as the system’s exclusive knowledge base. We conduct contextual inquiries with members of the academia and develop user experience (UX) artifacts, including personas and a consolidated sequence, to elicit information needs, identify breakdowns in the tasks, and derive system requirements. Our results indicate that UX research methodologies offer a robust and systematic foundation for characterising the design constraints required to tailor a generic RAG system to the needs of a specialised user community.

Index Terms—Artificial Intelligence, Retrieval-Augmented Generation, User Experience, Human-Centred Design

I. INTRODUCTION

The advent of LLMs profoundly transforms the field of information retrieval, with RAG [1] emerging as a leading approach combining generative capabilities with retrieval from external knowledge bases. While RAG proves effective for information needs [2], its adoption remains uneven, as studies show that traditional search engines are still preferred in generalist contexts, primarily due to interface familiarity [3]. This gap between technical promise and user acceptance underscores the need for UX-driven design research. A key question remains: *can RAG be adapted to serve specific user communities, and how should it be designed to meet their particular expectations?*

To address this question, we study the case of the *Biographie nationale de Belgique*, a scholarly biographical corpus (hereafter referred to as the corpus) currently available only in print or as PDF documents, intended for a broad audience ranging from academic researchers to anyone interested in the history of Belgium and its notable figures. In this study, we focus on expert users whose deep familiarity with the corpus makes them a particularly relevant population for eliciting

information needs and design requirements. This context offers a compelling case study for exploring how UX research methods inform the design of a RAG-based system tailored to a specific user community. Grounded in human-centred design, this work involves users throughout the design process to ensure that requirements emerge from situated use rather than abstract assumptions. We conduct contextual inquiries with four members of the academia to elicit their needs, identify recurring breakdowns, and derive design specifications for the proposed system. The remainder of this paper is organised as follows: Section II presents the theoretical background, Section III describes the methodology, Section IV reports the findings and introduces the proposed prototype, Section V discusses the research question, Section VI outlines the limitations and subsequent future work, and Section VIII concludes with contributions.

II. BACKGROUND

RAG architectures are increasingly used in information retrieval through the combination of document retrieval and LLM-based synthesis [1]. Prior work has explored domain-adapted RAG systems for specialized contexts such as biomedical research [4]. Similarly, digital humanities tools such as *Deutsche Biographie*¹ and *SampoSampo* [5] support access to biographical collections through semantic and generative approaches. However, these systems primarily focus on technical performance rather than users’ expectations and scholarly search practices. Research on information seeking behaviour [6] highlights the exploratory nature of academic search activities, emphasizing the importance of a user-centred perspective in the design of generative search tools.

UX is grounded in human-centred design, which positions users at the core of design and evaluation processes. Human-centred design aims to develop interactive systems by focusing on users’ needs, contexts of use, and experiential outcomes throughout the design lifecycle [7]. Within this framework, UX refers to “a person’s perceptions and responses that result from the use and/or anticipated use of a product, system or service,” encompassing emotional, cognitive, and behavioural dimensions before, during, and after interaction. Building on this foundation, our research operationalizes human-centred

The authors acknowledge the support of ILC and CENTAL of UCLouvain, as well as the support of *Académie royale de Belgique*. The authors also thank the four members of the *Commission de la Biographie* for their time, expertise, and valuable contributions to the study.

¹<https://www.deutsche-biographie.de>

design through two core principles that directly inform our methodological choices and empirical strategy [8].

- 1) The design is grounded in an explicit understanding of users, their tasks, and environments. In our case, this entails a close examination of a specific user group such as academicians who are also historians and of the epistemic context in which the system will be used. Academic research practices are shaped by disciplinary norms, standards of evidence, and interpretative workflows that structure how information is searched, evaluated, and integrated. Because these practices differ substantially from everyday web navigation, standard web usability conventions are insufficient to capture their requirements. Accordingly, interaction models and evaluation criteria are deliberately tailored to support domain-specific research activities rather than generic browsing behaviours [7].
- 2) Users are involved throughout the entire project lifecycle, from initial analysis to iterative refinement. During the analysis phase, we conduct contextual inquiries to identify users' situated practices, expectations, and constraints, ensuring that system requirements emerge from observed use rather than abstract assumptions. During design and development, we conduct formative evaluations through iterative design–evaluation cycles structured according to the UX Process Reference Model [9]. This framework provides a systematic articulation of analysis, design, prototyping, and evaluation activities, ensuring traceability between identified user requirements and successive design refinements. Empirical feedback collected at each stage informs successive refinements, enabling progressive alignment between user needs, system functionality, and experiential outcomes.

In addition to human-centred design, our approach is informed by the principles of human-centred artificial intelligence (HCAI) as articulated by [10]. HCAI emphasizes the development of AI systems that are reliable, safe, and trustworthy, while maintaining meaningful human control. Rather than positioning automation as a replacement for human judgment, this perspective frames AI as augmenting human capabilities, supporting oversight, transparency, and accountability. In our work, this orientation reinforces the commitment to designing systems that remain understandable and controllable by their users, particularly in contexts where epistemic responsibility and interpretative authority are central.

Finally, our framework incorporates the concept of trust in UX, a central construct in human–technology interaction. Among the various definitions proposed in the literature, we adopt the widely cited definition of [11], who describe trust as an attitude that an agent holds toward achieving a goal in situations characterized by uncertainty and vulnerability. Trust thus reflects a willingness to rely on a system despite incomplete knowledge and potential risk. In digital environments, this attitude plays a decisive role in adoption and sustained use: users are more likely to appropriate and integrate a system into

their practices when they perceive it as competent, predictable, and aligned with their goals. In our context, trust in the future system is not merely desirable but foundational to its acceptance, particularly given the epistemic stakes involved. Recent research, however, highlights an important nuance: trust and distrust are not simply opposite ends of a single continuum but distinct constructs that shape user experience and AI adoption in different ways [12]. While trust can promote reliance and engagement, distrust may independently influence emotional and behavioral responses, such as reluctance to disclose sensitive information or heightened scrutiny of system outputs. Evidence from studies on digital worker chatbots shows that distrust has its own explanatory power in predicting user reactions, beyond the absence of trust [12]. For our design, this implies that fostering adoption requires more than maximizing positive trust perceptions; it also demands proactively mitigating sources of distrust through transparency, predictability, and clear communication of system capabilities and limitations. Such an approach supports not blind reliance, but informed and sustainable acceptance.

III. METHODOLOGY

To conduct this research, we carry out a series of contextual inquiries, the rationale for which we outline below. We select participants through purposive sampling to reflect the disciplinary diversity. We transcribe all sessions to support the qualitative analysis (Section IV).

A. Contextual Inquiry

This research aims to understand users' information-seeking practices and recurring problems to design a trustworthy, user-tailored search engine. To this end, we opt for contextual inquiry [13], a qualitative method that involves observing and interviewing users in their natural work environment. Unlike other data collection methods such as questionnaires or focus groups, contextual inquiry allows researchers to access situated practices rather than post-hoc accounts or generalised opinions. This approach is particularly suited to our context: it enables participants to walk through their action sequences in detail, verbalise difficulties as they arise, and express frustrations in actual experience, thereby yielding rich, ecologically valid data. The data collection procedure is described in Subsection III-C.

B. Participant Sample

The *Académie royale de Belgique* is organised into four disciplinary classes (Sciences; Arts; Letters, Moral and Political Sciences; and Technology and Society), each comprising approximately four members. We interview one member per class, resulting in a purposive sample ($N=4$) that represents all four disciplinary classes (Table I). We conduct the contextual inquiry with these four members, whose mission is to document the lives and work of individuals who have achieved notable recognition in Belgium. We select this group for two complementary reasons. First, participants are experienced information seekers: writing biographical notices

requires them to mobilise a wide range of heterogeneous sources, which makes them a particularly relevant population for this study. Second, they constitute the primary target users of the proposed search engine. As historians, literary scholars, and scientists, they directly benefit from improved access to the corpus.

TABLE I
OVERVIEW OF THE PARTICIPANT PROFILES.

ID	Age	Discipline	Digital Proficiency
P1	65–70	Technology & Society	High
P2	80–90	Letters & Moral Sciences	Low
P3	80–90	Sciences	Medium
P4	80–90	Arts	Medium

C. Contextual Interviews

We conduct a total of eight interviews, two per participant. We organise the interviews in two stages. First, we hold an in-person warm-up session to establish initial contact and better understand each participant’s role within the academia. Before each session, we inform participants of the research objectives and obtain their verbal consent to record and use their data for research purposes. Second, we conduct four contextual interviews remotely via Zoom. Each participant joins from their home, office, or usual work environment. To elicit an authentic information-seeking sequence, we ask participants to reflect in advance on a biographical notice they recently wrote and to use it as a concrete anchor to retrace their process in detail, from identifying relevant sources to verifying and cross-checking information. This preparation helps maintain focus on situated information-seeking practices, resulting in sessions lasting approximately 15 minutes each.

D. Analytic Transcriptions

We audio-record the four contextual interviews (excluding the warm-up sessions) following participants’ informed verbal consent. We then produce a manual analytic retranscription of these sessions. Rather than generating verbatim transcripts, we selectively document action sequences, decision points, and salient insights that inform the sequence model, while removing conversational noise and off-topic exchanges. This retranscription process preserves the structural elements of participants’ information-seeking workflows and supports the qualitative analysis presented in Section IV.

IV. RESULTS

We analyse contextual data and produce four UX artifacts: two personas, a consolidated sequence model, a set of user requirement specifications, and a prototype that addresses the needs and constraints expressed by users.

A. Personas

The qualitative analysis reveals two distinct information-seeking strategies among participants. These differences relate not to disciplinary background or seniority, but to users’ relationships with digital tools. Based on this analysis, we

identify two analytical personas that inform the design of the proposed search engine:

- **The Hybrid User.** This profile adopts a combinatory approach to information seeking. The Hybrid User relies on the Internet in both a structured manner, using specialised databases such as Agatha or BelgicaPress, and in a more exploratory way when the direction of inquiry remains open. Digital research represents only one component of the overall workflow. Online investigations are systematically complemented by archival materials, institutional records, and human testimonies. As one participant explained, “*Sometimes we start with Google, then move back to archives or personal contacts.*” The research process involves continuous movement between digital, material, and interpersonal sources. A defining characteristic of this profile is a strong commitment to verification: digital results are treated as leads rather than validated facts, and cross-checking across multiple sources constitutes an essential step in the workflow.
- **The Selective Digital User.** This profile grounds information seeking primarily in established scholarly practices. The Selective Digital User relies first on personal expertise, prior research, and academic networks. As one participant described, research often begins with “*memories, previous works, books, or people who might be relevant to contact.*” Previously published biographies, particularly those from the *Biographie nationale*, serve as reliable starting points when available. Relevant institutions and organisations are consulted as complementary sources. The relationship to digital tools remains cautious and selective: search engine results are perceived as repetitive or insufficiently relevant, and AI-generated content is approached with skepticism. Frustration emerges with existing digital infrastructures, particularly when navigating PDF versions of the *Biographie*, which are perceived as cumbersome. In some cases, printing notices replaces on-screen navigation. Digital tools therefore function as complementary resources rather than central research environments.

These personas emerged from recurring attitudes toward digital tools observed across interviews rather than predefined categories. They are intended as analytical profiles supporting design reflection rather than generalized user typologies.

B. Consolidated Sequence

We identify two additional primary tasks during the elaboration of the individual sequences—namely, discussing the selection of biographical subjects and writing the notice itself but we do not include them in the present model, as this study focuses specifically on the information-seeking task. Although the two personas reflect distinct orientations toward digital tools, both engage in the same overarching task: searching for reliable biographical information. The consolidated sequence (Table II) shows that the search process unfolds systematically in four abstract stages, regardless of the individual strategy adopted. Breakdowns concentrate primarily in the first stage

(search online), where participants encounter unreliable, repetitive, or technically inaccessible content. We identify only one alternative strategy, relying on personal memory or previous work, which corresponds directly to the Selective Digital user profile and reflects a broader distrust of digital sources.

C. User Requirements Specifications

The UX artifacts produced here serve as the basis for deriving a set of user requirements. These requirements are organised into three categories : **functional**, **UX** and **trust-related** (Table III). Functional requirements define what the system must do: retrieving and synthesising information from the corpus, and exposing both a confidence score and the source passages alongside each response. UX requirements specify how the interaction should feel: delivering an immediate, synthesised answer that requires no advanced digital literacy, with fully copyable content. Finally, trust requirements address a broader concern expressed by participants: the reliability of information retrieved online, whether from search engines perceived as repetitive and sometimes inaccurate or from AI-generated content. All information must therefore be traceable to a verified source, and the system must clearly communicate that its knowledge base is restricted exclusively to the corpus.

D. Proposed System Design

RAG is an architecture that combines a retrieval component which searches a knowledge base for passages relevant to a query with a generative language model that synthesises a response from the retrieved content [1] (Figure 1). In this work, we adopt a generative approach while constraining the knowledge base exclusively to the corpus. Unlike systems that operate over open and uncontrolled sources, the proposed system ensures that generation is conditioned solely on verified content, rather than on parametric memory. This addresses a concern shared by both user profiles: the unreliability of information retrieved online, whether from repetitive search engine results or erroneous content.

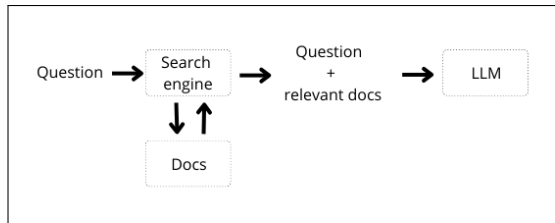


Fig. 1. Pipeline of RAG

As detailed in Table III, this architecture further addresses the specific needs of each profile. The system first performs a targeted retrieval over the corpus (**F1**), ensuring that all generated responses are grounded exclusively in this verified source (**T1**). For the *Hybrid* user, a confidence score (**F4**, **T3**) and the retrieved source passages (**F3**, **T2**) are exposed alongside each response, supporting his verification workflow.

For the *Selective Digital* user, a synthesised (**F2**) and immediate answer (**UX1**) replaces the conventional list of links, reducing cognitive load and requiring no advanced digital literacy (**UX3**). Additionally, all content is fully selectable and exportable (**UX2**). The interface therefore provides an export feature allowing users to download the generated response as a plain text or CSV file, directly addressing the copy-paste breakdown identified during the contextual inquiries. Figure 2 presents a mock-up of the proposed interface, with Claude (Anthropic) used as a design assistance tool during the prototyping process.

V. DISCUSSION

General-purpose RAG-based systems such as ChatGPT or Perplexity operate over open, uncontrolled corpora, returning results from heterogeneous sources of varying authority. For instance, querying Perplexity about RAG itself may return links to commercial websites or informal blogs, whereas an expert user would expect a reference to the foundational work [1]. While this is well-suited for everyday information needs, it falls short in specialised contexts where source reliability and verifiability are paramount.

By constraining the knowledge base exclusively to the corpus, the proposed system transforms a generic tool into a domain-specific one. This constraint simultaneously addresses the three categories of requirements identified in this study: it ensures functional reliability through a verified corpus, improves the user experience by eliminating irrelevant or repetitive results, and fosters trust by ensuring that all information is traceable to an authoritative source. This suggests that corpus control may function as a key design lever for adapting RAG to specialised scholarly communities such as the one examined in this study. While domain-adapted RAG systems are typically designed to improve performance in specialised domains, our work instead examines how retrieval grounded in a curated corpus can support scholarly practices and trust requirements. Unlike fine-tuning approaches, RAG preserves explicit links to source documents, supporting verification and provenance.

More broadly, this work suggests that UX research methods can serve as a principled basis for justifying technical design decisions in human-AI systems. Rather than selecting an architecture for technical convenience, our design choices are directly grounded in observed user needs and breakdowns, ensuring that the proposed system is both technically sound and human-centred.

VI. LIMITATIONS

This study remains at a design-oriented stage. The proposed system currently takes the form of a conceptual architecture and a high-fidelity mock-up, and has not yet been implemented as a fully functional RAG system. Consequently, we do not report empirical usability or adoption results at this stage. This limitation is consistent with a human-centred design approach, which recommends iterative cycles of analysis, prototyping, and evaluation prior to full-scale development [7], [9]. In line

TABLE II
CONSOLIDATED SEQUENCE MODEL FOR THE TASK “SEARCHING FOR RELIABLE INFORMATION,” DERIVED FROM FOUR CONTEXTUAL INQUIRIES AND SYNTHESISING RECURRING ACTION STEPS, BREAKDOWNS, AND ALTERNATIVE STRATEGIES ACROSS PARTICIPANTS.

Primary Task	Trigger	Abstract Steps	Breakdowns	Alternative Strategy
Search for reliable information	Decision to write a biographical notice	1. Search online	⚡ No information available online ⚡ Erroneous or contradictory results ⚡ AI-generated content unreliable ⚡ Website hard to use (e.g. no copy-paste)	<i>Rely on personal memory or prior research works</i>
		2. Consult documentary sources	—	
		3. Mobilise a human network	⚡ Approximate or unverifiable information	
		4. Cross-check and verify	—	

TABLE III
USER REQUIREMENTS SPECIFICATIONS

Category	ID	Requirement	Source
Functional	F1	Search and retrieve information from the corpus	Consolidated sequence — Step 1
	F2	Generate a synthesised response from the most relevant retrieved passages	Selective Digital persona
	F3	Display the source passages used to generate the response	Hybrid persona
	F4	Expose a confidence score alongside each response	Hybrid persona
UX	UX1	Deliver an immediate answer without requiring navigation across multiple links	Selective Digital persona
	UX2	Ensure all content is selectable and copyable	Selective Digital persona
	UX3	Require no advanced digital literacy to operate	Consolidated sequence
Trust	T1	Clearly indicate that all data originates exclusively from the <i>Biographie</i>	Both personas
	T2	Ensure all information is traceable to a verified source	Hybrid persona
	T3	Display the confidence score in a visible and interpretable manner	Hybrid persona

with this framework, the mock-up serves as an intermediate artefact designed to validate interaction principles, trust mechanisms, and requirement alignment before implementation.

In addition, the participant sample (N=4) reflects the disciplinary diversity of the academia, but does not capture the full range of potential user profiles beyond expert members. Subsequent research should extend testing to broader user groups, including less specialised audiences, in order to assess the transferability of the identified design principles.

VII. FUTURE WORK

Future work will focus on conducting formative usability evaluations of the mock-up with members of the academia. These evaluations will examine perceived usefulness, usability, trust calibration, and alignment with scholarly workflows. Insights will inform successive refinements before proceeding to the full development and deployment of the RAG system. This staged progression, from mock-up validation to later implementation, is consistent with our second human-centred design principle, which emphasises continuous user involvement and iterative refinement through design–evaluation cycles prior to full system development [7], [9]. Particular attention will be paid to whether the identified personas achieve comparable levels of usability and satisfaction when interacting with the system.

VIII. CONCLUSION

This work advances UX methodologies for AI-enabled systems, specifically in the context of RAG. Rather than treating RAG as a purely technical architecture, our findings suggest that contextual inquiry, persona modelling, and task sequence consolidation can systematically inform the design of domain-specific human–AI interaction. By grounding architectural choices in observed user practices, breakdowns, and trust concerns, this study positions UX research as a methodological bridge between AI capabilities and meaningful human experience. In doing so, it contributes to ongoing discussions on usability, trust, and design frameworks for specialised human–AI systems.

This study addresses the following research question: *can RAG be adapted to serve specific user communities, and how should it be designed to meet their particular expectations?* Through the elaboration of UX artifacts such as personas, consolidated sequence model, user requirements, and prototype, our findings suggest that RAG may be meaningfully adapted to the specific needs of a specialised scholarly community, while remaining exploratory and context-dependent. Beyond this specific case, this work highlights a broader issue: as RAG has become a dominant paradigm in information retrieval, researchers and practitioners tend to adopt it by default, with-

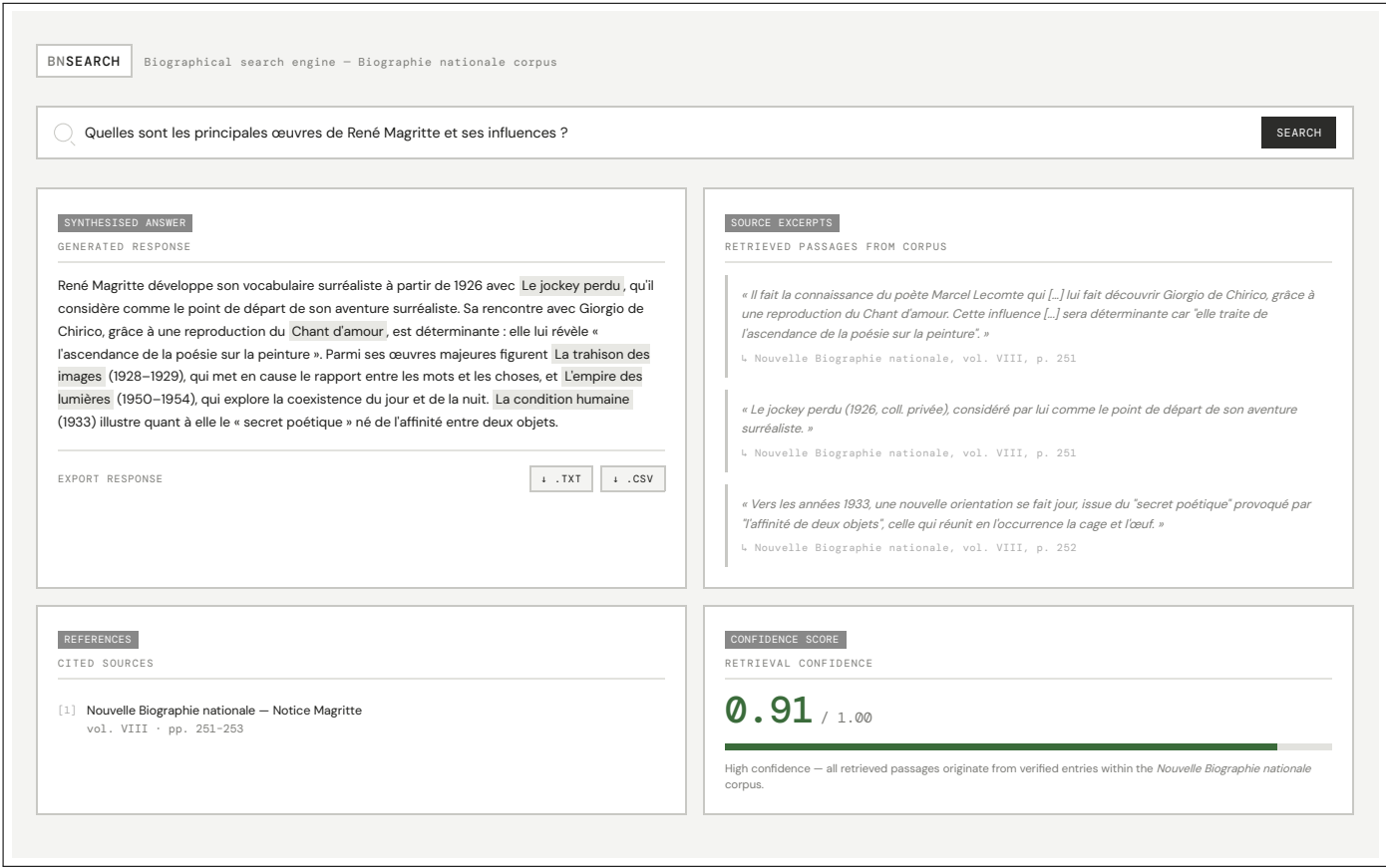


Fig. 2. Mock-up of the proposed interface, generated with the assistance of Claude (Anthropic).

out questioning how it should be tailored to the needs of its users. We argue that adapting RAG to specialised communities requires reframing it not as a purely technical architecture, but as a human–AI interaction system whose design must emerge from situated practices, epistemic norms, and trust dynamics.

REFERENCES

[1] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, “Retrieval-augmented generation for knowledge-intensive NLP tasks,” in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459–9474.

[2] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, and H. Wang, “Retrieval-augmented generation for large language models: A survey,” 2024. [Online]. Available: <https://arxiv.org/abs/2312.10997>

[3] S. E. A. Buananta, M. F. Hasani, A. N. Bara, and H. Wijaya, “User experience analysis when finding information: Comparative study of rag and traditional search engine,” in *2024 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. IEEE, 2024, pp. 798–803.

[4] C. Wang, Q. Long, M. Xiao, X. Cai, C. Wu, Z. Meng, X. Wang, and Y. Zhou, “Biorag: A rag-llm framework for biological question reasoning,” 2024. [Online]. Available: <https://arxiv.org/abs/2408.01107>

[5] A. Ahola, P. Leskinen, H. Rantala, J. Tuominen, and E. Hyvönen, “Using large language models for searching explainable relations in a cloud of cultural heritage knowledge graphs: Samposampo as a neuro-symbolic system,” in *DHNB 2026 Proceedings, Digital Humanities in the Nordic and Baltic Countries Publications*, vol. 8, no. 1. University of Oslo Library, Norway, 2026. [Online]. Available: <https://doi.org/10.5617/dhnbpub.13207>

[6] D. Ellis, “A behavioural approach to information retrieval system design,” *Journal of documentation*, vol. 45, no. 3, pp. 171–212, 1989.

[7] ISO 9241-210, “Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems,” International Organization for Standardization, Geneva, CH, Standard, 2019.

[8] M. Maguire, “Methods to support human-centred design,” *International journal of human-computer studies*, vol. 55, no. 4, pp. 587–634, 2001.

[9] S. Kieffer, L. Rukonić, V. Kervyn de Meerendré, and J. Vanderdonckt, “A process reference model for ux,” in *Computer Vision, Imaging and Computer Graphics Theory and Applications*, ser. Communications in Computer and Information Science, A. P. Cláudio *et al.*, Eds. Cham: Springer, 2020, vol. 1182, pp. 123–144.

[10] B. Shneiderman, “Human-centered artificial intelligence: Reliable, safe & trustworthy,” *International Journal of Human-Computer Interaction*, vol. 36, no. 6, pp. 495–504, 2020.

[11] J. D. Lee and K. A. See, “Trust in automation: Designing for appropriate reliance,” *Human factors*, vol. 46, no. 1, pp. 50–80, 2004.

[12] D. Gefen, B. A. Leffew, A. Ragowsky, and R. Riedl, “The importance of distrust in trusting digital worker chatbots,” *Communications of the ACM*, vol. 68, no. 4, pp. 42–49, 2025. [Online]. Available: <https://doi.org/10.1145/3701297>

[13] K. Holtzblatt, J. B. Wendell, and S. Wood, *Rapid contextual design: a how-to guide to key techniques for user-centered design*. Elsevier, 2004.