Interfaces to Encourage Critical User Engagement:
A Prototype Using RAG

Petra Dadić   Liana Ermakova

HCTI, Université de Bretagne Occidentale

dadic@univ-brest.fr, liana.ermakova@univ-brest.fr

Objective:

To support users in verifying AI-generated content by providing transparent source attribution in a RAG-enhanced interface.

Background:

LLMs like ChatGPT can produce hallucinations (false or misleading information).
Users struggle to assess factuality and trust sources.
RAG combines AI with document retrieval but lacks user-focused UI design.

Abstract:

AI-based chatbots using Large Language Models (LLMs) have become central to daily life. Still,they sometimes generate misleading information, known as "hallucinations," that can be hard for users to detect. This demo paper presents a prototype interface designed to help users identify and verify critical information in AI-generated content. Within the Retrieval-Augmented Generation (RAG) setting, the interface highlights key information and provides real-time access to supporting or contradictory sources. We conducted a study with 80 participants to gather feedback and refine the design, focusing on improving information sourcing and user trust. This paper demonstrates how thoughtful interface design can guide users to spot misinformation and enhance the utility of LLMs in information retrieval.

Usability Study (n=80)

We developed a web-based chatbot simulating a RAG environment with source attribution, evidence highlighting, and trust scores. Participants (mainly 25–34, France/Croatia, master's degrees) completed pre/post surveys and interacted with scripted prompts (some incorrect). 71% found the UI easy to use; highlighting was most valued, trust scores got mixed feedback. Users wanted more scientific sources, clearer source-claim links, optional highlighting, message history, and a tutorial. Insights informed an improved prototype now using real RAG with Elsevier abstracts.

Final Interface Design
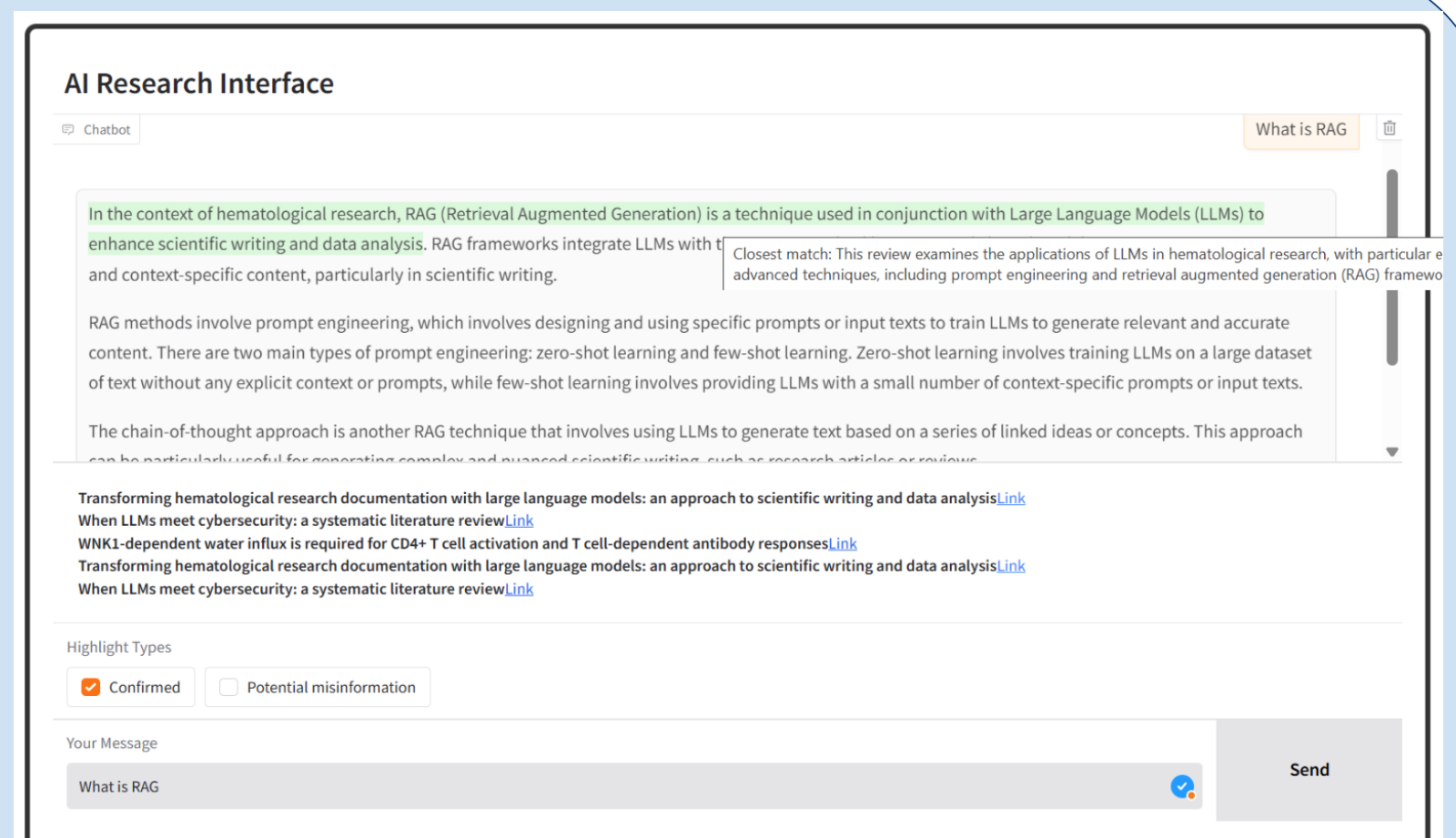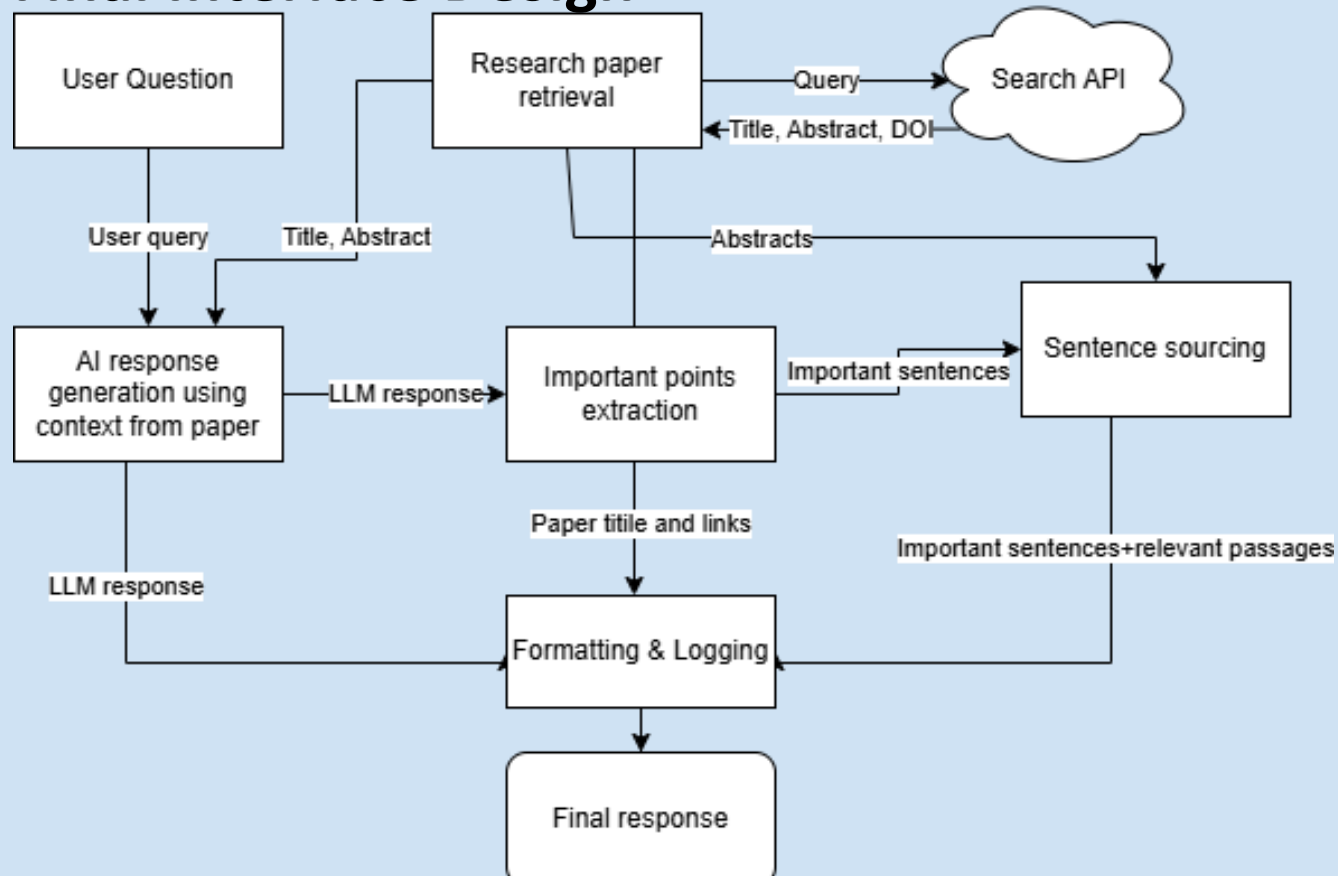
Evaluation

An initial evaluation (pending formal study) showed improved usability and trust; users preferred optional softer highlighting, no trust scores (relying on scientific sources only), clearer claim-to-source linking (e.g., tooltips), broader topic coverage, and interaction history. These findings guide ongoing refinements, with a formal study planned to assess impacts on transparency and data quality.

Conclusion

This demo presents a prototype supporting trustworthy, AI-assisted academic research through source retrieval, transparent attribution, and context-aware responses using Elsevier Scopus data. User feedback drove refinements like optional highlighting, interaction history, expanded topic coverage, and a softer color scheme. Trust scores got mixed reactions, shifting focus toward clearer source linking instead of numerical trust metrics. Current limitations include limited topics, reliance on abstracts, and model constraints; future improvements will add full-text integration, better citation accuracy, onboarding tutorials, and multilingual sources. This work underscores the value of user-centered AI for more transparent, accessible, and credible research tools.

Acknowledgements

Presented at CORIA 2025 – Authors: Petra Dadić and Liana Ermakova
Prototype: https://drive.google.com/file/d/1F_IGuI_QRocUql6Eor_BqS2oEQlDrAkA/view