

DIÁRIO DE PESQUISA ASSISTIDO POR IA COM LLMS ABERTAS E RAG

uma proposta de facilitação cognitiva para o trabalho científico cotidiano

AI-ASSISTED RESEARCH DIARY WITH OPEN LLMS AND RAG

A Cognitive Facilitation Approach for Everyday Scientific Work

Diego de Castro Rodrigues

Universidade de São Paulo, USP
Universidade Federal do Tocantins, UFT
E-mail: diegorodrigues.lua@gmail.com

Beatriz Rosa Pinheiro dos Santos

Universidade de São Paulo, USP
CeMEPI-PGE/SP
E-mail: bepinheiro@usp.br

Ildeberto Aparecido Rodello

Universidade de São Paulo, USP
E-mail: rodello@usp.br

Ieda Pelógia Martins Damian

Universidade de São Paulo, USP
Universidade Estadual Paulista, UNESP
E-mail: iedapm@usp.br

Luciana Romano Morilas

Universidade de São Paulo, USP
CeMEPI-PGE/SP
E-mail: morilas@fearp.usp.br

RESUMO

Este artigo apresenta uma arquitetura sociotécnica para diário de pesquisa assistido por inteligência artificial, voltada ao apoio do trabalho científico cotidiano. A proposta combina LLMS abertas com técnicas de recuperação de informação e geração aumentada por recuperação, incluindo RAG clássico, recuperação híbrida, recuperação temporal e mecanismos orientados por memória. O sistema foi concebido para operar sobre registros mínimos produzidos pelos pesquisadores e devolver perguntas organizadas, relações temáticas e sugestões de próximo passo sem ampliar a carga cognitiva do usuário. Para fundamentar a proposta, foi construído um conjunto sintético com 8 pesquisadores, 4 projetos, 8 áreas temáticas e 80 registros diários, preservando simultaneamente a perspectiva individual do pesquisador e a perspectiva agregada do centro de pesquisa. A

validação incidiu sobre três tarefas: extração de perguntas dos pesquisadores, identificação de convergências temáticas entre projetos e geração de estimativas de risco e entrega. Os resultados mostram que a estrutura de dados preserva sinais suficientes para diferenciar avanços, obstáculos e dúvidas, explicitar demandas metodológicas recorrentes, recuperar relações interdisciplinares e produzir indicadores úteis para acompanhamento em dois níveis. Conclui-se que a combinação entre LLMs abertas e RAG constitui uma base tecnicamente plausível para sistemas de facilitação cognitiva na pesquisa.

Palavras-chave: diário de pesquisa; LLMs abertas; RAG; recuperação de informação; facilitação cognitiva; processo científico.

ABSTRACT

This article presents a sociotechnical architecture for an AI-assisted research diary aimed at supporting everyday scientific work. The proposal combines open large language models (LLMs) with information retrieval and retrieval-augmented generation (RAG) techniques, including classic RAG, hybrid retrieval, temporal retrieval, and memory-oriented mechanisms. The system was designed to operate on minimal researcher inputs and to generate structured questions, thematic relationships, and next-step suggestions without increasing cognitive load. To support the proposal, a synthetic dataset was constructed with 8 researchers, 4 projects, 8 thematic areas, and 80 daily records, preserving both individual and collective research perspectives. The validation focused on three tasks: extracting researchers' questions, identifying thematic convergence across projects, and generating estimates of risk and delivery. The results show that the data structure preserves sufficient signals to distinguish progress, obstacles, and uncertainties, making it possible to identify recurring methodological demands, recover interdisciplinary relationships, and support monitoring at multiple levels. It is concluded that the combination of open LLMs and RAG constitutes a technically viable foundation for cognitive facilitation systems in scientific research.

Keywords: Research Diary; Open LLMs; RAG; Information Retrieval; Cognitive Facilitation; Scientific Process

REFERÊNCIAS

- ASAI, A.; ZHANG Z, W. Y. *et al.* Self-RAG: Learning to retrieve, generate, and critique through self-reflection. *In: The Twelfth International Conference on Learning Representations*, 2024.
- BENDER, E. M.; GEBRU T, M. M. A. *et al.* On the dangers of stochastic parrots: Can language models be too big? *In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp 610–623, 2021. DOI: <https://doi.org/10.1145/3442188.3445922>.
- EDGE, D.; TRINH H, C. N. *et al.* **From local to global:** A graph RAG approach to query-focused summarization. arXiv preprint arXiv:240416130, 2024. DOI: <https://doi.org/10.48550/arXiv.2404.16130>.
- GAO, Y.; XIONG Y, G. X. *et al.* **Retrieval-augmented generation for large language models:** A survey. arXiv preprint arXiv:231210997, 2023. DOI: <https://doi.org/10.48550/arXiv.2312.10997>.
- KARPUKHIN, V.; OĞUZ B, M. S. *et al.* Dense passage retrieval for open-domain question answering. *In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp 6769–6781, 2020. DOI: <https://doi.org/10.18653/v1/2020.emnlp-main.550>.
- LEWIS, P.; PEREZ E, P. A. *et al.* Retrieval-augmented generation for knowledge-intensive NLP tasks. *In: Advances in Neural Information Processing Systems*, pp 9459–9474, 2020.
- TOUVRON, H.; LAVRIL T, I. G. *et al.* **LLaMA:** Open and efficient foundation language models. arXiv preprint arXiv:230213971, 2023. DOI: <https://doi.org/10.48550/arXiv.2302.13971>.