

ChatFGV: Um protótipo de leitura de dados históricos via chat com LLMs

Isabela Yabe Martinez^b and George Dutra^b and Jaqueline Zulini^a and Felipe Vinhas^c and Fernanda Santos^c and Júlio César Chaves^b

^aCPDOC/Fundação Getúlio Vargas; ^bEMAp/Fundação Getúlio Vargas; ^cDTI/Fundação Getúlio Vargas

ARTICLE HISTORY

Compiled 15 de janeiro de 2026

Controle de versão - 15-janeiro - Apresentação para interessados no Latex.

Descrição da contribuição de cada autor: J.C. Chaves: idealização do projeto, supervisão geral, organização geral do artigo. Isabela Yabe Martinez: desenvolvimento da qualidade e validação do modelo, seção de validação e resultados. George Dutra: desenvolvimento do código para o RAG, experimentos com o DHBB, seção métodos. Jaqueline Zulini: apoio na idealização do projeto, revisão bibliográfica, seção de introdução e descrição do DHBB. Felipe Vinhas: apoio na infraestrutura de produção, revisão do artigo. Fernanda Santos: apoio na organização do projeto, revisão do artigo.

1. Introdução

1.1. Contexto e motivação

Em 1984, o Centro de Pesquisa e Documentação de História Contemporânea do Brasil da Fundação Getúlio Vargas (FGV CPDOC) publicou a primeira edição do Dicionário Histórico-Biográfico Brasileiro (DHBB). A obra se tornou a maior enciclopédia sobre a trajetória das elites políticas do país desde a década de 1930 até 1983, contendo quase 5 mil verbetes. Seu conteúdo incluía não somente os atores políticos que exerceram cargos de poder no nível federal (foco dos verbetes biográficos), mas também uma descrição de eventos históricos e instituições chave do Brasil de 1930 em diante, na

18 forma de verbetes temáticos.

19 2. Revisão Bibliográfica - inacabada

20 A partir de entrevistas com especialistas de mercado, Brotman and Sack (2025) e Ca-
21 deddu et al. (2024) destacam que o fenômeno transformador da IA é algo comparável
22 à energia elétrica no século XIX, com potencial para remodelar indústrias inteiras.
23 Carter (2025) enfatiza a importância de adotar uma mentalidade "AI First" para pros-
24 perar na era da superinteligência, destacando a necessidade de integrar a IA de forma
25 estratégica nos negócios, sem perder o toque humano.

26 3. O Dicionário Histórico-Biográfico Brasileiro (DHBB)

27 Idealizado por Israel Beloch, o DHBB foi concebido com a finalidade de sistematizar
28 informações até então dispersas sobre os principais atores e instituições da política
29 brasileira contemporânea. Inspirado por sua experiência prévia em projetos editoriais
30 e enciclopédicos, como a colaboração com Hélio Silva e a participação na Enciclopédia
31 Mirador, Beloch compreendeu, ainda na década de 1970, a necessidade de um ins-
32 trumento de consulta abrangente, factual e acessível para pesquisadores, estudantes e
33 demais interessados na história política do país Beloch (1978).

34 4. Métodos

35 Os verbetes do DHBB foram preparados para o RAG das LLMs através de um processo
36 básico de vetorização. Como os verbetes seguem um padrão bem estabelecido, não foi
37 necessário fazer modificações nem limpezas nos dados, apenas localizar os arquivos e
38 indicar os cabeçalhos como fonte dos meta-dados (formatados em YAML), [conforme](#)
39 [visto na seção de descrição do DHBB](#).

40 *Explicar o que é um system prompt.*

41 5. Validação

5.1. Coleta de dados

Para caracterizar o público-alvo e levantar necessidades reais de uso, foram aplicadas duas surveys online via Qualtrics. A Survey 1 contou com 34 respondentes, enquanto a Survey 2 contou com 74 respondentes. A coleta de dados foi conduzida em conformidade com os protocolos éticos aprovados pelo Comitê de Conformidade Ética em Pesquisas Envolvendo Seres Humanos da FGV (CEPH-FGV). O processo foi organizado em duas etapas: (i) um questionário inicial de perfil e necessidades e (ii) um questionário pós-interação com o protótipo.

6. Conclusão

- Contribuições
 - Recapitular como a combinação de LLMs com recuperação (RAG) reduz alucinações e melhora a precisão factual ao consultar o DHBB.
 - Enfatizar a importância de integrar e conectar fontes históricas para melhorar o processo de perguntas e respostas (QA).
- Limitações
 - Reconhecer restrições, como cobertura parcial do DHBB ou limitações de custo na hospedagem.
 - Mencionar o risco de indexação incompleta ou recuperação ruidosa.
 - Qualidade: realizar survey com perguntas e respostas humanas para avaliação.
- Direções Futuras
 - Sugerir aprimoramentos, como ampliar a cobertura para todo o DHBB, adotar modelos maiores ou mais especializados, ou incorporar algoritmos avançados de ranqueamento.
 - Indicar a possibilidade de adicionar abordagens de grafos de conhecimento ou técnicas avançadas de desambiguação.

- 69 I. Beloch. Dicionário histórico-biográfico brasileiro: dilemas na elaboração de um dici-
70 onário de história política. *Revista de Ciência Política*, 21(3):29–52, jul./set. 1978.
- 71 Adam Brotman and Andy Sack. *AI First: The Playbook for a Future-Proof Business*
72 *and Brand (English Edition)*. Harvard Business Review Press, 2025.
- 73 Andrea Cadeddu, Alessandro Chessa, Vincenzo De Leo, Gianni Fenu, Enrico Motta,
74 Francesco Osborne, Diego Reforgiato Recupero, Angelo Salatino, and Luca Secchi.
75 A comparative analysis of knowledge injection strategies for large language models
76 in the scholarly domain. *Engineering Applications of Artificial Intelligence*, 133, 7
77 2024. ISSN 09521976. .
- 78 Sandy Carter. *AI first, human always : embracing a new mindset for the era of supe-*
79 *rintelligence*. John Wiley & Sons, Inc., 2025. ISBN 1394189826.