

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 - 19-dez - Transcrita do Word para o LaTeX, a partir da versão do Word de 19-dez, última edição feita pela Isabela.

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

18 descrição de eventos históricos e instituições chave do Brasil de 1930 em diante, na
19 forma de verbetes temáticos.

20 **2. Revisão Bibliográfica - inacabada**

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

27 **3. O Dicionário Histórico-Biográfico Brasileiro (DHBB)**

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

35 **4. Métodos**

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

41 *Explicar o que é um system prompt.*

5. Validação	42
5.1. Coleta de dados	43
Para caracterizar o público-alvo e levantar necessidades reais de uso, foram aplicadas	44
duas surveys online via Qualtrics. A Survey 1 contou com 34 respondentes , enquanto a	45
Survey 2 contou com 74 respondentes. A coleta de dados foi conduzida em conformidade	46
com os protocolos éticos aprovados pelo Comitê de Conformidade Ética em Pesquisas	47
Envolvendo Seres Humanos da FGV (CEPH-FGV). O processo foi organizado em duas	48
etapas: (i) um questionário inicial de perfil e necessidades e (ii) um questionário pós-	49
interação com o protótipo.	50
6. Conclusão	51
• Contribuições	52
o Recapitular como a combinação de LLMs com recuperação (RAG) reduz	53
alucinações e melhora a precisão factual ao consultar o DHBB.	54
o Enfatizar a importância de integrar e conectar fontes históricas para melho-	55
rar o processo de perguntas e respostas (QA).	56
• Limitações	57
o Reconhecer restrições, como cobertura parcial do DHBB ou limitações de	58
custo na hospedagem.	59
o Mencionar o risco de indexação incompleta ou recuperação ruidosa.	60
o Qualidade: realizar survey com perguntas e respostas humanas para avalia-	61
ção.	62
• Direções Futuras	63
o Sugerir aprimoramentos, como ampliar a cobertura para todo o DHBB,	64
adotar modelos maiores ou mais especializados, ou incorporar algoritmos	65
avancados de ranqueamento.	66
o Indicar a possibilidade de adicionar abordagens de grafos de conhecimento	67
ou técnicas avançadas de desambiguação.	68

69 Referências

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