

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

Isabela Yabe Martinez<sup>b</sup> and George Dutra<sup>b</sup> and Jaqueline Zulini<sup>a</sup> and Felipe Vinhas<sup>c</sup> and Fernanda Santos<sup>c</sup> and Júlio César Chaves<sup>b</sup>

<sup>a</sup>CPDOC/Fundação Getúlio Vargas; <sup>b</sup>EMAp/Fundação Getúlio Vargas; <sup>c</sup>DTI/Fundação Getúlio Vargas

## ARTICLE HISTORY

Compiled 15 de janeiro de 2026

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

1

*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.

2  
3  
4  
5  
6  
7  
8

## 1. Introdução

9

### 1.1. Contexto e motivação

10

Em 1984, o Centro de Pesquisa e Documentação de História Contemporânea do Brasil da Fundação Getulio 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

11  
12  
13  
14  
15  
16  
17

<sup>18</sup> forma de verbetes temáticos.

## <sup>19</sup> 2. Revisão Bibliográfica - inacabada

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

## <sup>26</sup> 3. O Dicionário Histórico-Biográfico Brasileiro (DHBB)

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

## <sup>34</sup> 4. Métodos

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

<sup>40</sup> *Explicar o que é um system prompt.*

<b>41</b>	<b>5. Validação</b>	
	<b>5.1. Coleta de dados</b>	<b>42</b>
	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.	<b>43</b> <b>44</b> <b>45</b> <b>46</b> <b>47</b> <b>48</b> <b>49</b>
	<b>6. Conclusão</b>	<b>50</b>
	● Contribuições	<b>51</b>
	○ Recapitular como a combinação de LLMs com recuperação (RAG) reduz alucinações e melhora a precisão factual ao consultar o DHBB.	<b>52</b> <b>53</b>
	○ Enfatizar a importância de integrar e conectar fontes históricas para melhorar o processo de perguntas e respostas (QA).	<b>54</b> <b>55</b>
	● Limitações	<b>56</b>
	○ Reconhecer restrições, como cobertura parcial do DHBB ou limitações de custo na hospedagem.	<b>57</b> <b>58</b>
	○ Mencionar o risco de indexação incompleta ou recuperação ruidosa.	<b>59</b>
	○ Qualidade: realizar survey com perguntas e respostas humanas para avaliação.	<b>60</b> <b>61</b>
	● Direções Futuras	<b>62</b>
	○ Sugerir aprimoramentos, como ampliar a cobertura para todo o DHBB, adotar modelos maiores ou mais especializados, ou incorporar algoritmos avançados de ranqueamento.	<b>63</b> <b>64</b> <b>65</b>
	○ Indicar a possibilidade de adicionar abordagens de grafos de conhecimento ou técnicas avançadas de desambiguação.	<b>66</b> <b>67</b>

- 69 I. Beloch. Dicionário histórico-biográfico brasileiro: dilemas na elaboração de um dicio-  
70 nário 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.