

DANDARA.IA: Proposta de Chatbot Educacional com Fine-Tuning Culturalmente Responsivo para Educação Quilombola

Katherine Rojas Salazar
UNILA

Foz do Iguaçu, Brasil
soukatherinesalazar@gmail.com

Robert Isaías Agüero Bogado
UNILA

Foz do Iguaçu, Brasil
aguerorobert24@gmail.com

Abstract—This paper presents DANDARA.IA, a generative AI chatbot under development for quilombola school teachers in Brazil. Named after Dandara dos Palmares, the system aims to support educators in creating culturally contextualized educational content through LLM fine-tuning. We describe the technical architecture (Python backend, ReactJS frontend, GPT-4o fine-tuning), the proposed corpus curation methodology for Afro-Brazilian history and quilombola pedagogy, and the planned validation protocol. As a work in progress, we critically discuss technical challenges including: corpus diversity and bias, fine-tuning optimization for educational contexts, offline functionality for connectivity-limited regions, and alignment with National Curricular Guidelines for Quilombola Education (Resolution CNE/CEB 8/2012). This work contributes a replicable framework for developing culturally responsive educational AI while acknowledging current limitations and outlining validation roadmap.

Keywords—generative AI; educational chatbot; quilombola education; LLM fine-tuning; open source.

Resumo—Este artigo apresenta a DANDARA.IA, chatbot de IA generativa em desenvolvimento para professores de escolas quilombolas no Brasil. Nomeado em homenagem a Dandara dos Palmares, o sistema visa apoiar educadores na criação de conteúdo educacional culturalmente contextualizado através de fine-tuning de LLM. Descrevemos a arquitetura técnica (backend Python, frontend ReactJS, fine-tuning GPT-4o), a metodologia proposta de curadoria de corpus sobre história afro-brasileira e pedagogia quilombola, e o protocolo de validação planejado. Como trabalho em andamento, discutimos criticamente desafios técnicos incluindo: diversidade e viés do corpus, otimização de fine-tuning para contextos educacionais, funcionalidade offline para regiões com conectividade limitada, e alinhamento com as Diretrizes Curriculares Nacionais para Educação Quilombola (Resolução CNE/CEB 8/2012).

Palavras-chave—IA generativa; chatbot educacional; educação quilombola; fine-tuning de LLM; código aberto.

I. INTRODUÇÃO

Comunidades quilombolas brasileiras representam territórios de resistência histórica com 2.483 escolas atendendo 267.913 estudantes [1]. A Resolução CNE/CEB 8/2012 estabelece Diretrizes Curriculares Nacionais para Educação Escolar Quilombola [2], exigindo pedagogia contextualizada que integre memória coletiva, territorialidade e conhecimentos tradicionais. Contudo, 78% das escolas carecem de materiais específicos [3], e 88% dos professores não receberam formação adequada [4].

Paralelamente, modelos de linguagem de grande escala (LLMs) demonstram capacidades de personalização educacional [5], mas críticas apontam reprodução de vieses eurocêtricos [6] e inadequação para epistemologias não-hegemônicas [7]. Para comunidades tradicionais, tecnologias devem ser desenvolvidas com sensibilidade cultural e participação comunitária.

Este trabalho apresenta a **DANDARA.IA**, chatbot de IA generativa *em desenvolvimento* para professores quilombolas, homenageando Dandara dos Palmares. Como *work in progress*, descrevemos: (i) arquitetura técnica implementada; (ii) metodologia proposta de fine-tuning; (iii) plano de validação empírica; (iv) discussão crítica de limitações e desafios.

Contribuições: Framework técnico-metodológico para IA educacional contextualizada, análise de alinhamento com DCNs quilombolas, e roadmap transparente de desenvolvimento e validação para comunidades marginalizadas.

II. FUNDAMENTAÇÃO TEÓRICA

A. Educação Quilombola e DCNs

As DCNs quilombolas (Art. 7º, Res. 8/2012) estabelecem três eixos pedagógicos [2]: (i) memória coletiva e identidade étnica; (ii) línguas remanescentes e marcos civilizatórios;

(iii) territorialidade e práticas culturais. Silva [9] argumenta que pedagogia quilombola deve articular ancestralidade, território e resistência, não apenas “incluir” conteúdos afro-brasileiros. Santos e Nunes [10] demonstram que descolonização curricular exige processo coletivo com protagonismo comunitário.

A Lei 10.639/2003 [11] torna obrigatório ensino de história afro-brasileira, mas Gomes [12] evidencia lacuna entre marco legal e práticas escolares por escassez de recursos pedagógicos contextualizados.

B. IA Generativa: Críticas e Potenciais

LLMs podem personalizar educação [5], mas Bender et al. [6] alertam para “papagaios estocásticos” que reproduzem vieses sem compreensão. Noble [8] e Benjamin [7] demonstram como algoritmos perpetuam opressões raciais.

Watters [13] critica “solucionismo tecnológico” educacional, enquanto hooks [14] propõe pedagogia libertadora que centra agência docente. Para contextos tradicionais, tecnologias devem ampliar repertórios sem substituir autonomia [15].

C. Fine-tuning para Contextos Culturais

Fine-tuning adapta LLMs pré-treinados para domínios específicos via aprendizado supervisionado [16]. Trabalhos recentes demonstram viabilidade para línguas minoritárias, evidenciando potencial técnico com ressalvas éticas sobre homogeneização cultural.

Não identificamos sistemas validados para educação quilombola brasileira que articulem rigor técnico com participação comunitária – lacuna que este projeto busca preencher.

III. ARQUITETURA E METODOLOGIA

A. Visão Geral do Sistema

DANDARA.IA é chatbot web que permite professores gerarem: (i) planos de aula contextualizados; (ii) narrativas baseadas em tradições quilombolas; (iii) simulações interativas; (iv) atividades avaliativas alinhadas com DCNs. A Figura 1 apresenta a interface do sistema.

B. Stack Tecnológico

Backend: Python 3.11 com Flask para API REST. Autenticação via JWT, rate limiting (100 req/h), logs estruturados. Integração com API OpenAI para inferência do modelo fine-tuned.

Frontend: React 18 com design responsivo, acessibilidade WCAG 2.1 AA, otimizado para baixa banda. Interface com modos: (i) Online – sincroniza com servidor; (ii)

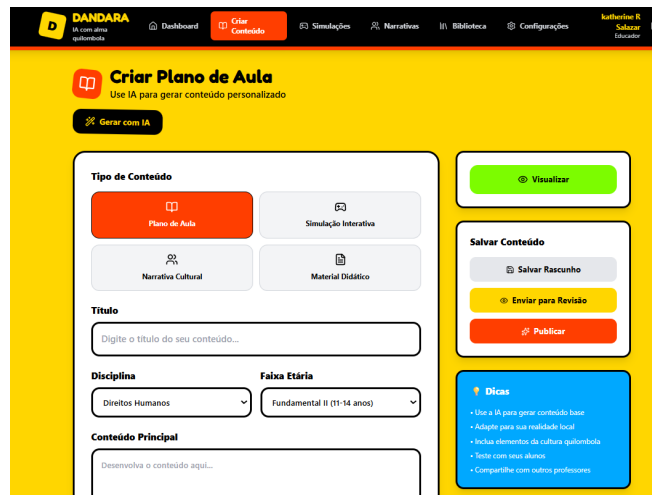


Figura 1. Interface da DANDARA.IA – Dashboard do educador com opções de criação de conteúdo personalizado para educação quilombola.

Híbrido – cache local + sync; (iii) Offline planejado – modelo compactado local.

Persistência: PostgreSQL 15 para metadados, histórico de gerações, feedback docente. Redis 7 para cache de respostas frequentes (reduzir latência e custos).

Infraestrutura: Deploy inicial em AWS EC2 t3.medium. Roadmap inclui migração para modelos open-source (LLaMA 3, Falcon) para soberania tecnológica.

C. Metodologia de Fine-tuning (Proposta)

Fase 1 – Curadoria de Corpus: Coleta sistemática de documentos sobre: (i) história afro-brasileira (fontes acadêmicas, IBGE, IPHAN); (ii) pedagogia quilombola (teses, dissertações, materiais do Programa Brasil Quilombola); (iii) marcos legais (DCNs, BNCC, Lei 10.639); (iv) narrativas orais transcritas (com consentimento comunitário). Meta: 8-10M tokens.

Crítérios de Seleção: Relevância educacional, precisão histórica verificável, alinhamento DCNs, diversidade regional. **Exclusão:** materiais com imprecisões, vieses discriminatórios, origem não-verificável.

Validação por Especialistas: Comitê consultivo (lideranças quilombolas + pesquisadores) avaliará amostra de 20% quanto a adequação cultural. Meta: concordância interavaliadores >0.80 (Kappa Cohen).

Pré-processamento: (i) Limpeza e normalização UTF-8; (ii) segmentação em unidades 2048 tokens; (iii) formatação instrução-resposta para supervised fine-tuning; (iv) anonimização de dados sensíveis; (v) marcação de metadados (região, nível, disciplina).

Fase 2 – Fine-tuning: Utilizaremos API OpenAI com GPT-4o base. Hiperparâmetros iniciais: learning rate 1.8e-5, batch size 8, 3 épocas, divisão treino/validação 90/10. Métrica primária: perda de validação. Análise qualitativa: 100 gerações avaliadas por especialistas quanto a adequação cultural e precisão pedagógica.

Fase 3 – Prompt Engineering: Sistema de prompts estruturados incorporando: (i) contexto das DCNs (Art. 7º); (ii) competências BNCC; (iii) parâmetros culturais (território, tradições); (iv) nível educacional e objetivos de aprendizagem.

D. Plano de Validação Empírica

Participantes: Professores de 5–8 escolas quilombolas (n20–30), critérios: >2 anos docência quilombola, formação pedagógica. Amostragem intencional buscando diversidade regional.

Protocolo Planejado – Fase Piloto (3 meses):

- Testes de usabilidade: cada professor gera 3 tipos de conteúdo, avalia via escala Likert (1-5) em dimensões: adequação cultural, relevância pedagógica, precisão histórica, alinhamento DCNs, usabilidade
- Entrevistas semiestruturadas (n=10-12, 45-60min): percepções, limitações, comparação com recursos disponíveis
- Análise de artefatos: conteúdos gerados avaliados por 3 especialistas independentes

Análise: Métodos mistos – estatísticas descritivas, concordância interavaliadores (Kappa), análise temática qualitativa.

Aspectos Éticos: Submissão a Comitê de Ética, consentimento informado, consulta prévia conforme Convenção 169 OIT, garantia de anonimato, direito de retirada.

IV. ALINHAMENTO COM DCNs QUILOMBOLAS

A Tabela I mapeia funcionalidades planejadas aos eixos das DCNs.

Estratégias de Implementação:

- Memória Coletiva:** Corpus incluirá narrativas históricas de resistência quilombola (Palmares, Calunga, Serra da Barriga). Sistema priorizará perspectivas afrocentradas
- Línguas Reminiscentes:** Fine-tuning incorporará vocabulário de matriz africana com significados contextualizados
- Territorialidade:** Prompts solicitarão conexão entre conteúdos curriculares e práticas territoriais quilombolas

TABELA I

ALINHAMENTO DANDARA.IA COM DCNs (RES. 8/2012)

Eixo DCN (Art. 7º)	Funcionalidade DANDARA.IA
I – Memória coletiva, identidade étnica	Geração de narrativas baseadas em histórias quilombolas específicas
II – Línguas remi- niscen- tes, marcos civiliza- tórios	Reconhecimento e uso con- textual de termos de matriz africana
III – Territorialidade, práticas culturais	Planos de aula integrando agricultura, artesanato, medicina tradicional

V. DESAFIOS E LIMITAÇÕES

A. Desafios Técnicos

Viés e Diversidade do Corpus: Risco de sub-representação regional (dados mais abundantes Sudeste/Sul vs. Norte/Nordeste). *Mitigação:* Curadoria intencional buscando equilíbrio, validação por especialistas de múltiplas regiões.

Otimização de Fine-tuning: Hiperparâmetros ótimos para domínio educacional quilombola são desconhecidos. *Abordagem:* Grid search em subset validação, análise qualitativa iterativa.

Funcionalidade Offline: 40% escolas quilombolas têm conectividade precária [17]. Modelos LLM são pesados (>10GB). *Roadmap:* (i) Implementar cache agressivo; (ii) investigar quantização/destilação de modelos; (iii) parcerias para infraestrutura comunitária.

Custos Computacionais: Fine-tuning GPT-4o: USD 500; inferência: USD 0.06/requisição. Escala nacional (2.483 escolas) exige recursos significativos. *Sustentabilidade:* Buscar financiamento via editais públicos, migração para modelos open-source.

B. Desafios Éticos e Sociais

Colonialidade Digital: Uso de infraestrutura OpenAI (EUA) processa dados quilombolas em servidores estrangeiros. *Reflexão Crítica:* Tensão entre viabilidade técnica atual e soberania de dados futura.

Risco de Homogeneização: IA pode padronizar práticas pedagógicas diversas. *Princípio:* Ferramenta como “amplificador de repertório”, não substituto de autonomia docente.

Sustentabilidade do Projeto: Risco de abandono pós-desenvolvimento inicial. *Compromisso:* Código open-source (GPL v3), documentação técnica completa, formação de comunidade de desenvolvedores.

VI. ESTADO ATUAL E PRÓXIMOS PASSOS

A. Estado Atual (Outubro 2025)

- **Arquitetura implementada:** Backend Python + Frontend React funcional
- **Protótipo funcional:** Testes internos demonstram viabilidade técnica
- **Corpus em curadoria:** Coletados 3.2M tokens (meta: 8-10M)
- **Comitê consultivo:** Em formação (2 lideranças confirmadas)
- **Fine-tuning:** Aguardando corpus completo
- **Validação empírica:** Planejada para início 2025

B. Roadmap de Desenvolvimento

Curto Prazo (3-6 meses): Completar curadoria de corpus, executar fine-tuning, conduzir validação piloto com 20-30 professores, publicar resultados preliminares.

Médio Prazo (6-12 meses): Refinar modelo baseado em feedback, implementar modo offline/híbrido, expandir para outras regiões quilombolas, desenvolver API REST para integração.

Longo Prazo (>12 meses): Migração para modelos open-source, adaptação para outras comunidades tradicionais, estudos longitudinais de impacto, construção de governança comunitária de dados.

validação, e à comunidade open-source que viabiliza tecnologias acessíveis.

REFERÊNCIAS

- [1] Instituto Nacional de Estudos e Pesquisas Educacionais, “Censo Escolar 2023,” INEP, Brasília, 2023.
- [2] Brasil, CNE/CEB, “Resolução nº 8: DCNs para Educação Escolar Quilombola,” Brasília, 2012.
- [3] J. M. Arruti, “Políticas públicas para quilombos,” in *Caminhos Convergentes*, IPEA, 2017.
- [4] S. A. Miranda, “Formação de professores quilombolas,” *Rev. Bras. Educação*, vol. 25, 2020.
- [5] E. Kasneci et al., “ChatGPT for good?,” *Learn. Individ. Differ.*, vol. 103, 2023.
- [6] E. M. Bender et al., “Stochastic parrots,” in *Proc. FAccT’21*, 2021.
- [7] R. Benjamin, *Race After Technology*. Polity, 2019.
- [8] S. U. Noble, *Algorithms of Oppression*. NYU Press, 2018.
- [9] D. M. Silva, “Educação quilombola,” *Cad. Pesquisa*, vol. 42, no. 146, 2012.
- [10] E. F. Santos and G. H. L. Nunes, “Currículo quilombola,” *Currículo sem Front.*, vol. 19, no. 3, 2019.
- [11] Brasil, “Lei 10.639/2003,” *Diário Oficial*, 2003.
- [12] N. L. Gomes, “Relações étnico-raciais,” *Currículo sem Front.*, vol. 12, no. 1, 2012.
- [13] A. Watters, *Teaching Machines*. MIT Press, 2021.
- [14] b. hooks, *Ensinando a Transgredir*. WMF Martins Fontes, 2013.
- [15] P. Freire, *Pedagogia do Oprimido*. Paz e Terra, 1970.
- [16] T. Brown et al., “Few-shot learners,” in *Proc. NeurIPS’20*, 2020.
- [17] CETIC.br, “TIC Educação 2022,” São Paulo, 2022.

VII. CONCLUSÃO

Este trabalho apresentou DANDARA.IA, chatbot educacional de IA generativa em desenvolvimento para professores quilombolas. Como *work in progress*, contribuímos com: (i) arquitetura técnica open-source implementada; (ii) metodologia proposta de fine-tuning culturalmente responsivo; (iii) plano de validação empírica rigoroso; (iv) análise de alinhamento com DCNs quilombolas; (v) discussão transparente de limitações e desafios éticos.

Reconhecemos estar em estágio inicial: corpus incompleto, modelo não fine-tuned, validação não realizada. Porém, ao compartilhar processo e desafios, buscamos: (i) feedback da comunidade científica; (ii) potenciais colaborações; (iii) contribuir para debate sobre IA educacional em contextos marginalizados.

Nosso compromisso é desenvolvimento participativo, com protagonismo quilombola, rigor metodológico, transparência sobre limitações, e código aberto. Tecnologias não são neutras – devem servir à justiça social e valorização de epistemologias historicamente silenciadas.

AGRADECIMENTOS

Às lideranças quilombolas que têm dialogado sobre o projeto, aos professores que manifestaram interesse na