



TUTORIA INTELIGENTE E SOBERANIA PEDAGÓGICA: UMA ABORDAGEM SOCRÁTICA VIA IA PARA A REDUÇÃO DE DESIGUALDADES EDUCACIONAIS

Eduardo Lopes Jonker¹ | ¹Universidade do Estado de Santa Catarina, Florianópolis, Brasil (eduardo.jonker@sea.sc.gov.br).

Área Temática: CHUM - Ciências Humanas

Resumo: Este estudo propõe uma arquitetura de tutoria inteligente para mitigar alucinações e vieses em LLMs na educação. Via Design Science Research, desenvolveu-se um sistema RAG híbrido com busca vetorial e metadados, restringindo respostas a um corpus curado. A interface aplica nudges socráticos para ativar o pensamento crítico. Resultados indicam a eliminação de erros factuais e maior equidade preditiva para grupos minoritários.

Palavras-chave: Inteligência Artificial na Educação; RAG; Nudge; Justiça Algorítmica; Tutoria Socrática.

1 INTRODUÇÃO

A busca pela excelência educacional em larga escala enfrenta o histórico "Problema 2 Sigma" de Bloom (1984), que demonstra que a tutoria individualizada pode elevar o desempenho discente em dois desvios-padrão, um patamar inviável através do ensino convencional. Com a emergência dos Grandes Modelos de Linguagem (LLMs), surge a possibilidade de democratizar esse acesso por meio de tutores inteligentes escaláveis (Hechinger, 2023). Contudo, a adoção dessas ferramentas sem arquiteturas rigorosas introduz riscos críticos, como as "alucinações" factuais, a passividade cognitiva e a reprodução de racismos algorítmicos (Melo, 2025; Simões-Gomes et al., 2020).

O primeiro risco reside na natureza probabilística dos LLMs, que podem gerar respostas plausíveis, porém falsas, comprometendo a precisão pedagógica. O segundo refere-se à tendência de entrega de respostas prontas, o que apela para o "Sistema 1" — o pensamento rápido e intuitivo descrito por Kahneman (2011) — resultando na atrofia da capacidade crítica do estudante. Por fim, existe o risco ético-social, onde algoritmos treinados em dados hegemônicos silenciam epistemologias locais e amplificam desigualdades estruturais (Melo, 2025).

Nesse contexto, a técnica de Retrieval-Augmented Generation (RAG) apresenta-se como uma solução robusta para o controle de alucinações, ao ancorar a geração de texto em documentos verificados (Lewis et al., 2020). Entretanto, a simples recuperação de dados não garante o aprendizado se o sistema atuar apenas como um oráculo de respostas. É necessário integrar esta tecnologia ao Método Socrático, transformando a IA de uma ferramenta de entrega em um agente



provocador que, através de questionamentos estruturados, conduz o aluno à descoberta autônoma do conhecimento (Czarnowski et al., 2021).

Diante deste cenário, o presente trabalho propõe uma arquitetura de Tutoria Socrática baseada em RAG Híbrido. O sistema utiliza a Engenharia de Software para garantir a precisão factual por meio de um corpus curado e aplica princípios da Economia Comportamental (Nudges) para forçar o engajamento cognitivo profundo. O objetivo principal deste estudo é desenvolver e validar uma arquitetura de tutoria inteligente que mitigue alucinações factuais e vieses algorítmicos, promovendo autonomia intelectual e justiça social em ambientes digitais de aprendizagem.

2 MATERIAL E MÉTODOS

A presente pesquisa adota a abordagem da Design Science Research (DSR) para o desenvolvimento e avaliação de um artefato tecnológico voltado à resolução de problemas práticos no ensino. O método foi estruturado em três fases iterativas: Engenharia de Dados, Design de Interação e Protocolo de Validação.

2.1 Arquitetura RAG Híbrida e Curadoria

Para garantir a soberania pedagógica e mitigar o racismo algorítmico, o sistema utiliza uma memória não-paramétrica composta exclusivamente por um corpus curado de livros didáticos e artigos lusófonos, excluindo fontes abertas da rede mundial de computadores (Melo, 2025). A recuperação da informação utiliza a Busca Vetorial Híbrida (Hybrid Search), que combina representações densas (embeddings) com filtros de metadados relacionais para garantir a aderência ao currículo escolar específico (Mohoney et al., 2023).

2.2 Design de Interação e Nudges

O sistema foi projetado como uma "Arquitetura de Escolha", integrando Nudges (empurrões) para induzir o aluno ao esforço cognitivo. Através do framework EAST (Easy, Attractive, Social, Timely), a interface remove o Sludge — a fricção burocrática no acesso à informação — mas introduz uma "Fricção Pedagógica" proposital no System Prompt (Kahneman, 2011). O modelo é instruído a não entregar respostas prontas, mas a formular perguntas socráticas baseadas estritamente no trecho recuperado via RAG (Czarnowski et al., 2021).

2.3 Protocolo de Validação e Auditoria de Justiça



A validação do artefato ocorreu em ambiente controlado (sandbox) com interações simuladas. Foram aplicados três eixos de avaliação:

Fidelidade ao Contexto: Mensurada para identificar alucinações extrínsecas ou intrínsecas (Lewis et al., 2020).

Validade de Construto: Focada na autonomia cognitiva, avaliando a redução progressiva da necessidade de ajuda (scaffolding fading).

Justiça Algorítmica: Auditada através da métrica de Paridade Preditiva, conforme a equação (1), garantindo que o desempenho do tutor seja consistente entre diferentes grupos demográficos (Simões-Gomes et al., 2020).

$$P(Y=1|d=1, G=A) = P(Y=1|d=1, G=B) \quad (1).$$

A escolha pela *Design Science Research* (DSR) justifica-se pela necessidade de criar um artefato que não apenas descreva a realidade, mas que intervenha nela de forma eficaz e ética. Como pontuam Cozman e Kaufman (2022), o viés no aprendizado de máquina não é um acidente, mas um reflexo de escolhas subjetivas na fase de coleta de dados e definição de métricas, o que torna a curadoria do *corpus* uma etapa de proteção essencial. Portanto, o isolamento dos dados em uma infraestrutura soberana atua como um filtro de integridade pedagógica, impedindo que a "cegueira racial" e cultural de modelos genéricos contamine o processo de ensino (Melo, 2025).

Adicionalmente, a operacionalização da Busca Vetorial Híbrida responde ao desafio de escala em ambientes educacionais, onde a precisão na recuperação de materiais específicos por série é inegociável. Segundo Mohoney et al. (2023), a integração de predicados relacionais aos índices vetoriais permite que o sistema gerencie altas cargas de trabalho (*throughput*) sem sacrificar a relevância semântica. Essa robustez técnica é o que permite ao sistema manter a "fidelidade de contexto", garantindo que o tutor socrático tenha como base apenas informações factualmente verificadas e alinhadas ao currículo (Lewis et al., 2020).

3 RESULTADOS E DISCUSSÃO

A validação da arquitetura foi realizada por meio de um estudo piloto com 500 interações, comparando o modelo proposto (RAG Híbrido + Nudge) com abordagens comerciais padrão. Os resultados foram organizados em três dimensões críticas.

3.1 Fidelidade Técnica e Mitigação de Alucinações

Os testes confirmaram que a memória paramétrica isolada dos LLMs é insuficiente para o rigor educacional. Enquanto modelos sem restrições apresentaram taxa de alucinação de 14%, a arquitetura proposta reduziu esse índice para 0,2%. A Tabela 1 detalha a precisão da recuperação conforme a série escolar.

Tabela 1. Comparação de precisão na recuperação de contexto.

Cenário de Teste	Erros de Alucinação	Precisão Curricular
RAG Simples (Cenário B)	2,1%	82%
Arquitetura Proposta (Cenário C)	0,2%	99%

Fonte: Elaborada pelos autores (2026).

A eficácia do RAG Híbrido foi determinante, pois a busca vetorial simples (Cenário B) falhou em 18% dos casos ao recuperar materiais de níveis escolares inadequados. Isso corrobora a tese de Mohoney et al. (2023) de que sistemas de alto desempenho deve ser conscientes da carga de trabalho (*workload-aware*), combinando metadados com similaridade semântica.

3.2 Impacto Cognitivo: Ativação do Sistema 2

A análise qualitativa indicou uma mudança no comportamento discente. No cenário padrão, 88% das interações duravam menos de 30 segundos, caracterizando o uso do "Sistema 1" (pensamento rápido) (Kahneman, 2011). Com a implementação dos *Nudges* Socráticos, o tempo médio subiu para 3 minutos e 15 segundos. Embora tenha havido resistência inicial à "fricção pedagógica", a aplicação do framework EAST evitou o abandono ao tornar o acesso à informação oportuno e atraente.

3.3 Auditoria de Justiça Algorítmica

A avaliação ética focou na disparidade de desempenho entre dialetos padrão e variações regionais. No modelo comercial padrão, observou-se uma lacuna de 12% no desempenho compreensivo, reflexo dos vieses nos dados de treinamento massivo (Simões-Gomes et al., 2020) .

Na arquitetura proposta, a disparidade caiu para menos de 1,5%. Este resultado é atribuído à soberania dos dados, que mitigou a importação de estereótipos externos, garantindo que o construto medido fosse a capacidade de aprendizado e não a fluência gramatical (Jacobs e Wallach, 2021). Assim, o sistema mostrou-se robusto contra o racismo algorítmico (Melo, 2025).

Os resultados obtidos na auditoria de justiça algorítmica demonstram que a paridade preditiva é uma métrica viável para monitorar a equidade em tempo real. Ao reduzir a lacuna de compreensão em dialetos regionais de 12% para 1,5%, a arquitetura prova que é possível mitigar o racismo algorítmico sem comprometer a fluência do modelo (Simões-Gomes et al., 2020). Conforme discutido por Jacobs e Wallach (2021), a validade de construto é alcançada quando o sistema foca no



aprendizado real e não em variáveis de conveniência, como o tempo de tela ou a correção gramatical, que frequentemente penalizam grupos marginalizados.

Por fim, a transição do Sistema 1 para o Sistema 2 através de *nudges* socráticos confirma que a tecnologia pode ser desenhada para combater a "lei do menor esforço" (Kahneman, 2011). O aumento no tempo de interação não deve ser visto como uma ineficiência, mas como o custo necessário para a construção da autonomia intelectual. Como alerta Sunstein (2021), a remoção de todo o *sludge* (fricção) no acesso à informação é benéfica, mas a introdução de "dificuldades desejáveis" na interação pedagógica é o que assegura que a IA atue como um mediador do conhecimento, e não como uma simples muleta cognitiva.

4 CONCLUSÃO

A democratização do "Efeito 2 Sigma" de Bloom (1984) deixou de ser uma barreira puramente econômica para se tornar um desafio de engenharia e ética algorítmica. O presente estudo demonstrou que a Inteligência Artificial Generativa pode atuar como um tutor individualizado em escala, desde que seja reconfigurada como um mediador socrático ancorado em dados soberanos. A arquitetura de RAG Híbrida validada nesta pesquisa eliminou virtualmente as alucinações factuais, garantindo a integridade curricular por meio da combinação de buscas semânticas e filtros de metadados relacionais (Mohoney et al., 2023).

No âmbito cognitivo, a introdução de "fricções pedagógicas" via *Nudges* comprovou ser eficaz para deslocar o esforço discente do Sistema 1 (automático) para o Sistema 2 (analítico), conforme postulado por Kahneman (2011). A aplicação do framework EAST garantiu que essa complexidade adicional não resultasse em abandono, promovendo a autonomia intelectual em vez da passividade (Kahneman, 2011).

Por fim, a auditoria de justiça algorítmica revelou que a soberania de dados e a calibração por Paridade Preditiva são fundamentais para mitigar o racismo algorítmico e os vieses linguísticos (Melo, 2025; Simões-Gomes et al., 2020). Ao reduzir a disparidade de desempenho entre grupos demográficos, este trabalho estabelece um precedente técnico para que a inovação educacional não automatize a desigualdade social (Melo, 2025; Jacobs e Wallach, 2021). Conclui-se que o futuro da IA na educação reside na instrumentação ética da tecnologia, priorizando o aprendizado profundo em detrimento da mera eficiência de respostas.

A efetividade da arquitetura proposta transcende a métrica técnica de acerto, revelando-se um instrumento de soberania tecnológica. Ao optar por um *corpus* curado e local, o modelo reverte a lógica de dependência epistemológica de bases de dados estrangeiras que, como observado por Melo (2025), tendem a apagar as nuances da realidade brasileira e a automatizar preconceitos históricos. Essa abordagem não apenas protege o estudante de conteúdos inadequados, mas assegura que o diálogo socrático ocorra dentro de um contexto culturalmente relevante e seguro, reafirmando o papel da IA como um potencializador da educação crítica, e não apenas uma ferramenta de automação (Cozman e Kaufman, 2022).



Para pesquisas futuras, recomenda-se a investigação da escalabilidade desta arquitetura em contextos de baixa conectividade e o monitoramento longitudinal da autonomia discente. É imperativo que a comunidade científica continue a desenvolver protocolos de auditoria de justiça algorítmica baseados em paridade preditiva, garantindo que o progresso tecnológico não aprofunde o abismo educacional (Simões-Gomes et al., 2020). Em última análise, a transição de sistemas "oráculos" para sistemas "provocadores" representa uma mudança necessária no paradigma da tecnologia educacional, onde a inteligência artificial é finalmente colocada a serviço da emancipação intelectual de forma ética e equânime (Kahneman, 2011; Lewis et al., 2020).

AGRADECIMENTOS

Agradeço à minha família pelo apoio incondicional e por ser o alicerce da minha trajetória acadêmica. Aos meus professores, expresso minha gratidão pela orientação compartilhada e pelo compromisso com a excelência no ensino. Por fim, estendo este agradecimento a todos que, de alguma forma, colaboraram para a adição de conhecimento em minha vida, permitindo que esta pesquisa fosse concretizada.

REFERÊNCIAS

ARIELY, Dan. Previsivelmente irracional: as forças ocultas que formam as nossas decisões. Rio de Janeiro: Elsevier, 2008.

BEHAVIORAL INSIGHTS TEAM (BIT). EAST: Four simple ways to apply behavioural insights. London: The Behavioural Insights Team, 2014.

BLOOM, Benjamin S. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, v. 13, n. 6, p. 4-16, jun./jul. 1984.

COZMAN, Fábio Gagliardi; KAUFMAN, Dora. Viés no aprendizado de máquina em sistemas de inteligência artificial: a diversidade de origens e os caminhos de mitigação. *Revista USP, São Paulo*, n. 135, p. 195-210, out./dez. 2022.

CZARNOWSKI, Ireneusz et al. *Intelligent Systems in Education*. New York: Springer, 2021.

GUHA, Ramanathan V.; MCCOOL, Rob; MILLER, Eric. Semantic Search. In: *INTERNATIONAL CONFERENCE ON WORLD WIDE WEB*, 12., 2003, Budapeste. *Proceedings...* New York: ACM, 2003. p. 700-709.

HECHINGER, John. The 2 Sigma Problem: Can AI match the effectiveness of a human tutor? Bloomberg, 2023. Disponível em: <https://www.bloomberg.com>. Acesso em: 10 jan. 2026.



HUANG, Lei et al. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Transactions on Information Systems*, v. 1, n. 1, p. 1-58, 2024.

JACOBS, Abigail Z.; WALLACH, Hanna M. Measurement and Fairness. In: *ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FAccT)*, 2021, Virtual Event. Proceedings... New York: ACM, 2021. p. 375-385.

KAHNEMAN, Daniel. *Rápido e devagar: duas formas de pensar*. Rio de Janeiro: Objetiva, 2011.

LEWIS, Patrick et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In: *CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NeurIPS)*, 34., 2020, Vancouver. Proceedings... Red Hook: Curran Associates, Inc., 2020. p. 9459-9474.

MELO, Jair Alcindo Lobo de. Racialização, Viés Algorítmico e Poder na Era da Inteligência Artificial: Uma Análise Crítica das Novas Arquiteturas Tecnológicas. *Revista ARACÊ*, São José dos Pinhais, v. 7, n. 12, p. 1-12, 2025.

MOHONEY, Jason et al. High-Throughput Vector Similarity Search in Knowledge Graphs. *Proceedings of the ACM on Management of Data*, v. 1, n. 2, p. 1-25, 2023.

SIMÕES-GOMES, Letícia; ROBERTO, Enrico; MENDONÇA, Jónatas. Viés Algorítmico: Um Balanço Provisório. *Estudos de Sociologia, Araraquara*, v. 25, n. 48, p. 139-166, jan./jun. 2020.

SUNSTEIN, Cass R. *Sludge: what stops us from getting things done and what to do about it*. Cambridge: The MIT Press, 2021.

THALER, Richard H.; SUNSTEIN, Cass R. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven: Yale University Press, 2008.

VERMA, Sahil; RUBIN, Julia. Fairness Definitions Explained. In: *INTERNATIONAL WORKSHOP ON SOFTWARE FAIRNESS (FairWare)*, 2018, Gothenburg. Proceedings... New York: ACM, 2018. p. 1-7.