Projeto de P&D e Governança em IA Educacional: Protocolo de Formação Docente, Soberania de Dados e Transferência Tecnológica (LENFI/UFSCar - PNED/MEC)

Adriana Mattar Maamari

Crivum DA82 · DOI 10.67665/crivum.2x5g5hig · publicada em 23/09/2026

CASE DE INOVAÇÃO E PROJETO DE TECNOLOGIA EDUCACIONAL Título: Projeto de P&D e Governança em IA Educacional: Protocolo de Formação Docente, Soberania de Dados e Transferência Tecnológica (LENFI/UFSCar - PNED/MEC) Proponente e Liderança Científica: Profa. Dra. Adriana Mattar Maamari — Universidade Federal de São Carlos (UFSCar)

1. Enquadramento, Objeto de Inovação e Definições Operacionais

O avanço acelerado de Modelos de Linguagem de Grande Porte (LLMs) na Educação Básica exige soluções operacionais e metodológicas para a transformação digital das redes públicas de ensino. Em alinhamento estrito com a Política Nacional de Educação Digital (PNED - Lei nº 14.533/2023) e as diretrizes do Ministério da Educação (MEC), o objeto deste projeto é o desenvolvimento, a validação empírica e a transferência tecnológica de uma metodologia de governança e engenharia de prompts voltada à formação docente contínua, desenvolvida no âmbito do Laboratório de Ensino de Filosofia (LENFI/UFSCar).

1.1. Definições Operacionais e Mecanismo Cognitivo

Governança de IA Educacional: Conjunto integrado de protocolos pedagógicos, regulatórios e computacionais que normatizam a seleção, a auditabilidade conceitual, a privacidade de dados (LGPD) e o uso ético-crítico de LLMs no ambiente escolar, impedindo o consumo passivo de respostas prontas.

Auditoria Conceitual: Procedimento analítico sistemático de checagem de saídas geradas por LLMs, estruturado para a identificação de alucinações de dados, anacronismos históricos, falácias lógicas e vieses ideológicos subjacentes aos modelos de linguagem.

Mecanismo Cognitivo e Indicador Validado (IACA): A auditoria opera como um dispositivo disparador de dissonância cognitiva (Gallo, 2012). O ganho em competência argumentativa dos estudantes é mensurado pelo Instrumento de Avaliação da Coerência Argumentativa (IACA) — adaptado do modelo estrutural de Toulmin (1958) —, avaliando a capacidade do estudante de identificar premissas, garantias, refutações e modalizadores em fases pré-intervenção (T_0) e pós-intervenção (T_2).

1.2. Matriz Hermenêutico-Algorítmica e Camadas Operacionais

Diferente das rubricas padronizadas de AI Literacy (Ng et al., 2021), que avaliam predominantemente a usabilidade técnica e a acurácia factual genérica, a Matriz Hermenêutico-Algorítmica atua diretamente na estrutura lógico-conceitual do pensamento filosófico e crítico, dividindo-se em três camadas operacionais encadeadas:

Camada 1 (Sintética / Chain-of-Thought - T_0): Produção de síntese conceitual via encadeamento de raciocínio lógico pelo LLM (Wei et al., 2022), na qual o modelo é provocado a gerar um mapa inicial de conceitos com base no cânone filosófico.

Camada 2 (Auditoria Hermenêutica / Benchmarking - T_1): Avaliação cega da saída sintética por meio da aplicação da rubrica IACA pelos estudantes e docentes, pontuando a resposta em quatro dimensões filosóficas na escala Likert 0-3 (precisão conceitual, contextualização histórica, validade lógica e qualidade das fontes).

Camada 3 (Objeção Dialética / Counter-Prompting - T_2): Submissão do modelo a restrições de Self-Consistency (Wang et al., 2022), forçando o algoritmo a sustentar e justificar a antítese conceitual diante do texto de referência e das objeções formuladas pelos alunos.

2. Ancoragem em Referências, Estado da Arte e Mapeamento da Lacuna (PRISMA)

2.1. Mapeamento PRISMA e Tabela de Extração (n = 11)

A Revisão Sistemática da Literatura (2023-2026) foi conduzida nas bases SciELO, Scopus, ERIC, BDTD/CAPES, LA Referencia e Google Acadêmico, seguindo rigorosamente o protocolo PRISMA: Identificação inicial (n = 214) -> Removidos por duplicidade (n = 34) -> Triagem de título e resumo (n = 180) -> Excluídos por irrelevância ao escopo (n = 138) -> Avaliação do texto completo (n = 42) -> Amostra final sintetizada (n = 11).

A extração sistemática da amostra final (n = 11) revelou três grandes blocos temáticos no estado da arte:

Estudos focados em diretrizes genéricas e éticas de IA na educação (UNESCO, 2023; Holmes & Tuomi, 2022);

Trabalhos voltados a frameworks institucionais e operacionais de AI Literacy (Ng et al., 2021; Mollick & Mollick, 2023);

Artigos focados no uso de LLMs para disciplinas STEM (Ciência, Tecnologia, Engenharia e Matemática).

Conclusão da Extração: Ficou demonstrada a existência de uma lacuna crítica na literatura internacional: nenhum dos 11 estudos analisados desenvolve ou valida empiricamente protocolos de engenharia de prompts, governança e auditoria crítica de LLMs aplicados especificamente ao ensino de Ciências Humanas e Filosofia na Educação Básica sob arquiteturas de execução local.

2.2. Tensionamento Teórico da Literatura

A fundamentação teórica da proposta articula ativamente duas frentes de pesquisa frequentemente isoladas:

Engenharia de Prompting e Avaliação Computacional de LLMs: Incorpora os conceitos de Chain-of-Thought (Wei et al., 2022), Self-Consistency (Wang et al., 2022), mapeamento sistemático de alucinações (Ji et al., 2023) e alfabetização algorítmica (Ng et al., 2021).

Plataformização Crítica e Didática da Filosofia: Dialoga com a crítica à corporatização tecnológica da escola pública (Selwyn, 2019), com o uso pedagógico ativo de agentes de IA (Mollick & Mollick, 2023) e com as matrizes brasileiras de didática da experiência filosófica (Gallo, 2012; Maamari, 2020, 2024), convertendo a falha algorítmica e a alucinação em matéria-prima para o exercício do pensamento reflexivo.

2.3. Diretrizes Normativas e Arquitetura de Infraestrutura

A arquitetura tecnológica do projeto atende plenamente às exigências da Lei nº 14.533/2023 (PNED), da Portaria MEC nº 2.158/2023 e das diretrizes globais da UNESCO (2023):

Ambiente Off-line (Edge AI): Execução local do modelo aberto Llama-3-8B-Instruct quantizado (Q4_K_M) em contêineres pré-configurados Docker/Ollama.

Hardware Comercial de Baixo Custo: Operação completa em estações de trabalho convencionais (ex.: CPU AMD Ryzen 7 ou Intel Core i7, 16 GB RAM DDR4), dispensando placas gráficas dedicadas (GPUs) de alto custo.

Conformidade Estrita com a LGPD: Garantia de 100% de soberania de dados através da retenção local de prompts e textos escolares (arquitetura Zero Egress), impedindo o envio de produções de crianças e adolescentes para servidores externos ou corporativos.

3. Equipe Técnica e Ancoragem Institucional

A execução, a validação de campo e o acompanhamento pedagógico do projeto apoiam-se na infraestrutura consolidada de pesquisa, pós-graduação e extensão da Universidade Federal de São Carlos (UFSCar):

Laboratório de Ensino de Filosofia (LENFI/UFSCar): Atua como núcleo central de prototipagem, análise de mídias educacionais, engenharia de runtime e desenvolvimento de Recursos Educacionais Abertos (REA).

Programa de Pós-Graduação Profissional em Filosofia (Prof-Filo / Polo UFSCar): Assegura o canal direto de articulação com os professores regentes da rede pública estadual de ensino, viabilizando a aplicação da pesquisa-ação e a validação do protocolo em ambiente escolar real.

Liderança Científica (Profa. Dra. Adriana Mattar Maamari): Docente do Departamento de Filosofia da UFSCar, com vasta trajetória acadêmica e liderança de projetos de pesquisa focados em Didática da Filosofia, formação de professores, análise de políticas públicas educacionais e tecnologias no ensino.

4. Trajetória, Produção Científica e Ativos Tecnológicos

A qualificação da liderança e da equipe do LENFI/UFSCar traduz-se em uma sólida produção acadêmica e no desenvolvimento de ativos aplicados à educação pública:

Produção Científica Qualificada: Publicação contínua de artigos em periódicos qualificados, e-books e capítulos de livros (Maamari, 2020, 2024; Maamari & Carlos, 2024) focados na especificidade da Didática da Filosofia, formação docente e análise crítica de políticas curriculares.

Ativos Tecnológico-Educacionais: Desenvolvimento, catalogação e difusão de Recursos Educacionais Abertos (REA), e-books pedagógicos e guias de orientação docente registrados com ISBN, servindo de suporte direto para os cursos de extensão e para a rede de egressos do Prof-Filo.

5. Desenho Metodológico, Séries Temporais e Robustez Estatística

5.1. Delineamento Quase-Experimental com Séries Temporais Interrompidas (STI) Para isolar estatisticamente o efeito da Matriz Hermenêutico-Algorítmica em relação ao aprendizado geral decorrente do curso de capacitação, controlando ameaças de maturidade e história, adota-se um delineamento quase-experimental com Grupo Controle Não Equivalente e medições repetidas em quatro janelas temporais (T_0, T_1, T_2, T_3):

Grupo Intervenção (N_int = 30): Docentes regentes e suas respectivas turmas submetidos ao módulo formativo de 30 horas focado no uso da Matriz Hermenêutico-Algorítmica e na Auditoria Conceitual de LLMs.

Grupo Controle (N_ctrl = 30): Docentes regentes e suas respectivas turmas submetidos a módulo formativo de 30 horas de carga horária equivalente sobre uso de tecnologias digitais e IA genérica (ferramentas comerciais tradicionais de mercado), sem o protocolo específico de auditoria hermenêutica.

Poder Estatístico e Amostragem: O cálculo do poder amostral (1 - beta = 0,80, com alpha = 0,05 e tamanho de efeito médio esperado d de Cohen = 0,65) determinou a necessidade de uma amostra mínima de N = 52 participantes. A amostragem total foi fixada em N_total = 60 (30 no Grupo Intervenção e 30 no Grupo Controle) para absorver eventuais atritos e desistências ao longo do processo.

Medições Longitudinais (Séries Temporais Interrompidas - STI):

T_0 (Mês 3): Aplicação do diagnóstico de linha de base (pré-intervenção) utilizando o instrumento IACA em ambos os grupos.

T_1 (Mês 8): Medição intermediária realizada ao término do módulo de formação docente.

T_2 (Mês 14): Medição pós-intervenção realizada após seis meses de aplicação prática do protocolo em sala de aula pelos professores regentes.

T_3 (Mês 20): Medição residual longitudinal realizada seis meses após o término do acompanhamento direto, para avaliar a retenção de longo prazo e o impacto sustentado da inovação.

Plano de Análise Estatística: Aplicação de Análise de Variância de Medidas Repetidas (RM-ANOVA) para testar a interação estatística entre Grupo x Tempo, acompanhada de testes t de Student para amostras independentes, testes exatos de Fisher para variáveis categóricas de detecção de erros e cálculo das dimensões de efeito (d de Cohen e eta quadrado parcial).

Confiabilidade da Análise Qualitativa: A Análise de Conteúdo (Bardin, 2016) e a atribuição de pontuações IACA aos textos produzidos pelos estudantes serão conduzidas em procedimento duplo-cego por dois codificadores independentes. A concordância interavaliadores será validada pelo

Coeficiente Kappa de Cohen, exigindo-se valor mínimo de kappa >= 0,80 para garantia de consistência interna.

Procedimentos Éticos: O protocolo experimental cumpre integralmente as exigências da Comissão Nacional de Ética em Pesquisa (Resoluções CNS nº 466/2012 e 510/2016), estando submetido ao Comitê de Ética em Pesquisa da UFSCar (CEP/UFSCar via Plataforma Brasil), com garantia de Assentimento e Termo de Consentimento Livre e Esclarecido (TCLE).

5.2. Critérios Observáveis de Falseamento da Intervenção

Para assegurar o rigor científico e a transparência metodológica, a hipótese de eficácia da inovação será considerada refutada/falseada se qualquer uma das seguintes condições for observada ao final do ciclo experimental (T_2/T_3):

Insignificância do Efeito: A variação média no índice de coerência argumentativa (IACA) no Grupo Intervenção (N_int) não apresentar superioridade estatisticamente significativa ($p < 0,05$; $d >= 0,5$) em relação ao Grupo Controle (N_ctrl).

Inoperância da Auditoria: Persistência de alucinações conceituais e erros de lógica formal não detectados em mais de 20% das saídas auditadas pelos estudantes na Camada 2 da Matriz.

Rejeição da Usabilidade Tecnológica: Índice de percepção desfavorável ou rejeição técnica superior a 35% nos grupos focais docentes em relação à viabilidade de operação do ambiente off-line (Llama 3 / Docker) nas computadores das escolas.

5.3. Matriz de Gestão de Riscos Operacionais e Epistemológicos

Risco 1 (Atrição da Amostra Docente): Abandono ou desligamento de professores ao longo dos 20 meses de pesquisa-ação.

Mitigação: Oferta de certificação oficial de extensão universitária emitida pela Pró-Reitoria de Extensão (ProEx/UFSCar) proporcional aos módulos concluídos, aliada ao suporte pedagógico assíncrono.

Risco 2 (Restrições de Infraestrutura Escolar): Instabilidade de energia ou equipamentos obsoletos nas escolas estaduais.

Mitigação: Execução 100% off-line da arquitetura Docker/Ollama embarcada em pendrives bootáveis ou instalações locais em estações de trabalho comuns.

Risco 3 (Resistência ou Uso Acrítico da Tecnologia): Tendência de docentes ou alunos utilizarem o LLM como mero gerador de textos prontos.

Mitigação: Estruturação de um currículo formativo estritamente centrado na desconstrução do erro algorítmico e na refutação dialética.

6. Reprodutibilidade, Parâmetros Técnicos e Anexo IACA

6.1. Configuração Técnica e System Prompt Template

Ambiente de Execução: Ollama (v0.1.32) executado em contêiner isolado Docker (Ubuntu 22.04 LTS).

Modelo Base: Meta Llama-3-8B-Instruct quantizado em 4-bits (Q4_K_M).

Parâmetros de Inferência Fixados: Temperature = 0.2; Top_P = 0.9; Seed = 42; Context Window = 4096 tokens.

System Prompt Template (Camada 1 e 3):

"Você é um assistente acadêmico focado no ensino crítico de Filosofia. Sua função é responder à consulta do estudante apresentando a tese do autor, os argumentos centrais e, obrigatoriamente, levantar 2 objeções clássicas ou limites conceituais dessa posição, sem fornecer respostas dogmáticas prontas."

6.2. Estrutura do Anexo IACA (Instrumento de Avaliação da Coerência Argumentativa) O instrumento IACA avalia quatro dimensões da produção textual/auditada na escala Likert de 0 a 3:

Identificação da Tese e Premissas (0-3): Capacidade de isolar a afirmação central e os fundamentos lógicos do texto ou da resposta gerada pela IA.

Detecção de Garantias e Fontes (0-3): Verificação de que as passagens citadas correspondem ao contexto do autor do cânone e não a invenções algorítmicas.

Formulação de Objeções e Refutações (0-3): Habilidade de articular contraexemplos, apontar falácias ou contestar premissas frágeis.

Coerência e Modalização (0-3): Uso adequado de qualificadores modais e ausência de contradições lógicas internas no texto final.

7. Referências Bibliográficas

BARDIN, Laurence. Análise de conteúdo. São Paulo: Edições 70, 2016.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília: Presidência da República, 2018.

BRASIL. Lei nº 14.533, de 11 de janeiro de 2023. Institui a Política Nacional de Educação Digital (PNED). Brasília: Presidência da República, 2023.

BRASIL. Ministério da Educação. Portaria MEC nº 2.158, de 28 de dezembro de 2023: Estabelece diretrizes para a Educação Digital e uso de tecnologias nas redes públicas. Brasília: MEC, 2023.

BRASIL. Conselho Nacional de Saúde. Resolução CNS nº 466, de 12 de dezembro de 2012: Diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. Brasília: Ministério da Saúde, 2012.

BRASIL. Conselho Nacional de Saúde. Resolução CNS nº 510, de 07 de abril de 2016: Normas aplicáveis a pesquisas em Ciências Humanas e Sociais. Brasília: Ministério da Saúde, 2016.

DETTMERS, Tim et al. QLoRA: Efficient Finetuning of Quantized LLMs. Advances in Neural Information Processing Systems, v. 36, p. 10088-10115, 2023.

FRANTAR, Elias et al. GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers. arXiv preprint arXiv:2210.17323, 2022.

GALLO, Silvio. Metodologia do ensino de Filosofia: uma didática da experiência filosófica. Campinas: Papirus, 2012.

HOLMES, Wayne; TUOMI, Ilkka. State of the art and practice in AI in education. European Journal of Education, v. 57, n. 4, p. 542-570, 2022.

JI, Ziwei et al. Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, v. 55, n. 12, p. 1-38, 2023.

MAAMARI, Adriana Mattar. A especificidade e a transversalidade da Didática da Filosofia na formação de professores. Educação e Filosofia, v. 38, n. 1, p. 115-142, 2024.

MAAMARI, Adriana Mattar; CARLOS, Diogo da Silva. O ensino de filosofia com crianças e jovens: experiências do LENFI. São Carlos: Pedro & João Editores, 2024.

MOLLICK, Ethan R.; MOLLICK, Lilach. Assigning AI: Seven Approaches for Students, with Prompts. Wharton School Working Paper, p. 1-42, 2023.

NG, Davy T. K. et al. AI literacy: Definition, teaching, evaluation and ethical issues. Computers and Education: Artificial Intelligence, v. 2, p. 100033, 2021.

ROGERS, Everett M. Diffusion of Innovations. 5. ed. New York: Free Press, 2003.

SELWYN, Neil. Should Robots Replace Teachers? AI and the Future of Education. Cambridge: Polity Press, 2019.

THIOLLENT, Michel. Metodologia da pesquisa-ação. 18. ed. São Paulo: Cortez, 2011.

TOULMIN, Stephen E. The Uses of Argument. Cambridge: Cambridge University Press, 1958.

TRIPP, David. Pesquisa-ação: uma introdução metodológica. Educação e Pesquisa, v. 31, n. 3, p. 443-466, 2005.

UNESCO. Guidance for Generative AI in Education and Research. Paris: UNESCO Publishing, 2023.

WANG, Xuezhi et al. Self-Consistency Improves Chain of Thought Reasoning in Language Models. arXiv preprint arXiv:2203.11171, 2022.

WEI, Jason et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. Advances in Neural Information Processing Systems, v. 35, p. 24824-24837, 2022.

ZHENG, Lianmin et al. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. Advances in Neural Information Processing Systems, v. 36, p. 46595-46623, 2023.