

Desafios da representação científica local em Sistemas de Inteligência Artificial: alternativas decoloniais para a Educação Tecnológica

Luiz Batista de Oliveira Neto¹

Marcelo Sabbatini²

Universidade Federal de Pernambuco – UFPE

Resumo

A sub-representação do conhecimento científico do Sul Global em Inteligência Artificial (IA) constitui uma faceta do colonialismo de dados, conforme modelos de linguagem são treinados predominantemente com dados norte-americanos e europeus (70%), relegando o Sul Global a fornecedor de dados e consumidor de tecnologia e grandes corporações se apropriam de conhecimento local sem creditação, perpetuando estruturas coloniais. A partir de uma revisão da literatura, investigamos a possibilidade do uso de *Small Language Models* (SLMs), aliado a práticas alternativas decoloniais. Esses modelos permitiriam o treinamento de IAs com dados científicos locais e a preservação de epistemologias contextualizadas, a partir de ações de desenvolvimento técnico consciente do contexto, tutela reversa e solidariedade.

Palavra-chave: colonialismo de dados; soberania tecnológica; epistemologia decolonial; Sul Global.

Introdução

No século XXI, com a expansão da *internet* e o surgimento da *big data*,³ nossos dados pessoais se tornaram o combustível da engrenagem digital. Cada clique, busca e interação *online* gera informações valiosas que alimentam algoritmos e direcionam publicidades customizadas. O que antes estava restrito a bibliotecas físicas agora é totalmente digital, transformando como acessamos informações e, principalmente, como somos monitorados nesse processo.

As grandes corporações GAFAM (*Google, Apple, Facebook/Meta, Amazon e Microsoft*) extraem os dados e monitoram comportamentos constantemente da população, moldando as ações em benefício próprio (Silveira *et al.*, 2021; Zuboof, 2021). Silveira *et al.* (2021) classificam isso como colonialismo de dados, uma nova

¹ Licenciado em Educação Física, mestrando do Programa de Pós-Graduação em Educação Matemática e Tecnológica, Universidade Federal de Pernambuco – UFPE. E-mail: luiz.batistaoliveira@ufpe.br

² Doutor em Teoria e História da Educação, professor do Centro de Educação da Universidade Federal de Pernambuco – UFPE. E-mail: marcelo.sabbatini@ufpe.br

³ Segundo Russell e Norvig (2021), *big data* consiste em grandes volumes de dados que requerem sistemas computacionais especializados para armazenamento e processamento.

versão das “veias abertas” denunciadas por Eduardo Galeano em seu clássico: em vez de ouro e prata, agora são extraídos os dados pessoais.

Com as tecnologias digitais e grandes corporações, enfrentamos os desafios da Inteligência Artificial (IA), definida por Russell e Norvig (2021) como a ciência de tornar máquinas inteligentes. Essa área se popularizou com os grandes *chatbots*, que utilizam técnicas como *machine learning* (aprendizado de máquina) e *deep learning* (aprendizado profundo), mas carrega consigo um problema estrutural: seus algoritmos são treinados predominantemente com dados do Norte Global.

Contudo, a utilização massiva de sistemas de Inteligência Artificial na educação revela a sub-representação sistemática do conhecimento científico do Sul Global (Manvi *et al.*, 2024). Essa relação é evidenciada pela forma como grandes corporações tecnológicas se apropriam do conhecimento do Sul Global, sem creditação aos pesquisadores e comunidades originárias, configurando um colonialismo digital onde empresas extraem gratuitamente dados e saberes.

Apesar dos olhares estarem centralizados apenas aos pequenos modelos de Inteligência Artificial Generativa (IAG), esta concentração cria um cenário onde *small language models* (SLMs) emergem como alternativa viável para democratizar o acesso à IA educacional, permitindo treinamento com dados científicos locais e abordagens pedagógicas contextualizadas (Wang *et al.*, 2025). O desenvolvimento de estratégias decoloniais para IA educacional torna-se imperativo para países do Sul Global construírem soberania tecnológica e preservarem seus saberes científicos e culturais.

Diante da sistemática sub-representação do conhecimento científico do Sul Global nos sistemas de IA educacionais, que perpetua desigualdades históricas e limita o potencial tecnológico em contextos não hegemônicos, justifica-se investigar alternativas decoloniais para construir soberania algorítmica. Assim, os objetivos deste estudo são analisar como a concentração de dados do Norte Global reforça o colonialismo de dados e avaliar o potencial dos *Small Language Models* (SLMs) como estratégia para democratização da IA na educação.

Metodologia

Este estudo adota uma abordagem qualitativa, fundamentada em pesquisa bibliográfica. O objetivo central é analisar a sub-representação científica do Sul Global

em sistemas de IA e discutir alternativas decoloniais para a educação tecnológica, com base em referências teóricas sobre colonialismo de dados, soberania tecnológica e epistemologias decoloniais, articulando esses conceitos para compreender o problema proposto. O levantamento do material teórico foi realizado através de busca sistemática em bases de dados como *SciELO* e *Google Acadêmico*, além do repositório de *pre-prints* arXiv, para acessar pesquisas recentes. A estratégia de busca utilizou descritores específicos como “colonialismo de dados”, “soberania tecnológica”, “inteligência artificial na educação” e “*small language models*”, aplicados para mapear o campo e identificar as principais fontes sobre o tema.

A seleção das fontes que compõem o corpus final da análise seguiu critério de inclusão baseado na relevância temática. Priorizaram-se trabalhos que não apenas diagnosticassem o problema da sub-representação do Sul Global, mas que também analisassem ou propusessem alternativas decoloniais para a tecnologia educacional, incluindo estudos sobre SLMs como ferramentas potenciais para democratização do acesso à Inteligência Artificial, garantindo o alinhamento com os objetivos deste estudo.

O problema da sub-representação científica em sistemas de IA

A sub-representação científica de países periféricos evidencia a assimetria que ocorre nos treinamentos das inteligências artificiais, mantendo a hegemonia ideológica dos países nos quais são desenvolvidas. De acordo com os estudos de Manvi *et al.* (2024), os modelos de LLMs não são treinados com dados provenientes de países subdesenvolvidos, como aqueles do continente asiático, países do Sul Global e da África.

Com baixa representatividade científica dos países subdesenvolvidos na produção dessas tecnologias, observa-se que estes assumem o papel de meros utilizadores e fornecedores de dados para as grandes corporações. Configura-se assim o que Silveira *et al.* (2021) denomina de “colonialismo de dados”, no qual os países desenvolvidos têm como função extrair os bens mais preciosos da era digital: os dados pessoais de suas populações.

Em uma sociedade onde o controle dos dados é monopolizado por países anglo-saxões, evidencia-se como esses modelos reproduzem narrativas hegemônicas e controlam as perspectivas históricas dominantes. Essa concentração de poder

tecnológico resulta na padronização de visões de mundo que são disseminadas globalmente através das respostas dos *chatbots*.

Nesse sentido, reforça-se a necessidade de representação científica do continente asiático, países do Sul Global e da África, como ruptura epistêmica com as estruturas hegemônicas do conhecimento. Contudo, as grandes corporações digitais revelam a apropriação de dados culturais para treinar modelos de IA perpetuando o epistemicídio através da ocultação da autoria (Santos, 2017; Faustino; Lippold, 2023).

Educação tecnológica

A tecnologia sempre acompanhou o desenvolvimento humano, mas a revolução digital transformou as relações de comunicação e produção de conhecimento. Na educação brasileira, essa integração foi formalizada pelo Plano Nacional de Educação (PNE) de 2014, através das "práticas pedagógicas inovadoras" (Brasil, 2014). Contudo, foi a pandemia de COVID-19 que acelerou drasticamente essa transição, impondo o ensino remoto e expondo a dependência de plataformas tecnológicas externas.

As tecnologias digitais implementadas na educação brasileira são predominantemente desenvolvidas por países do Norte Global, configurando uma relação de dependência tecnológica estrutural. Essa dinâmica reproduz o que Marini (2000) caracteriza como capitalismo dependente, no qual países hegemônicos definem os papéis e funções dos países periféricos no sistema global, perpetuando assimetrias de poder que se estendem ao campo educacional.

Apesar da dependência tecnológica é importante considerar os estudos de Silva (2023), que destaca a necessidade urgente da inclusão digital na educação. O autor aponta que a ausência de letramento tecnológico tanto por parte dos alunos quanto dos professores representa um obstáculo significativo ao processo educacional. Nesse contexto, a falta de competências digitais pode ampliar as desigualdades no ambiente educacional, criando um cenário preocupante de exclusão digital que afeta diretamente o desenvolvimento acadêmico dos estudantes.

Contudo, é fundamental refletir sobre as potencialidades que as tecnologias digitais podem apresentar para a educação brasileira. Monte (2025), em seus estudos, explora essas potencialidades, argumentando que as tecnologias digitais, quando criticamente implementadas, podem apresentar características transformadoras no processo educacional. A autora destaca que essas ferramentas oferecem possibilidades

de democratização do acesso à informação, flexibilidade metodológica, aprendizagem colaborativa e personalização do ensino que considere as singularidades dos educandos.

No entanto, é importante ressaltar que a tecnologia, por si só, não é capaz de resolver os problemas estruturais da educação brasileira. Sua potencialidade só poderá ser alcançada por meio de uma articulação eficiente entre governo, instituições de ensino, professores, alunos e comunidade escolar. Somente com investimentos contínuos e progressivo conseguiremos mudar esse cenário.

Impactos específicos na educação tecnológica

As tecnologias digitais na educação passam por transformações significativas com a expansão da plataformização do ensino. Dijck e Poell (2018) classificam este fenômeno como a inserção de plataformas digitais na educação, englobando recursos audiovisuais, videoconferência, IA e redes sociais metrificando e datafizando todo ensino. Vieira e Ferraro (2025, p.13) afirmam que “as plataformas precarizam ainda mais o trabalho docente, pois podem diminuir a autonomia pedagógica dos professores ao transformá-los em gestores de dados”.

A plataformização educacional revela problemas de transparência e ética. O caso do Paraná exemplifica essa problemática através do reconhecimento facial escolar. Segundo Israel *et al.* (2023), essas tecnologias operavam sem transparência sobre o destino das imagens dos estudantes, capturando expressões faciais através do Educatron⁴. Esta prática configura-se como capitalismo de vigilância, onde a plataformas captam as informações sem o consentimento das pessoas⁵.

Assim, a inserção da Inteligência Artificial na educação deve ser analisada sob a perspectiva do colonialismo de dados. Embora essas tecnologias possam auxiliar no ensino, é fundamental questionar quem programa esses modelos e quais são os interesses comerciais, tendo em vista que os dados são terceirizados para as *bigtechs* e

⁴ Os Professores foram informados pela direção da escola que participariam de um experimento com a câmera do Kit Educatron composto por computador e câmera presentes em todas as salas de aula do estado. A câmera passaria a captar informações faciais de modo contínuo, durante as aulas, para medir o desempenho dos alunos, com o objetivo de gerar gráficos e índices sobre o rendimento da turma (Israel *et al.*, 2023, p. 28).

⁵ O capitalismo de vigilância constitui um modelo econômico baseado na apropriação e comercialização de dados pessoais, empregando essas informações para antecipar condutas e moldar escolhas individuais, frequentemente operando à margem da consciência e consentimento dos usuários (Zuboff, 2021).

estas detêm o controle das plataformas digitais que estão inseridas nas escolas (Silveira *et al.*, 2021).

A pesquisa realizada por Cruz *et al.*, (2024) revela as seguintes estatísticas sobre a situação brasileira: 74% dos domínios de e-mails de instituições de Ensino superior são armazenados no *Google* e 9% na *Microsoft*. Este cenário demonstra como “mal conseguimos manter os dados das universidades nas próprias universidades. Os dados da educação pública brasileira são usados para treinar a inteligência artificial dessas grandes empresas”.

Paradoxalmente, enquanto se discute o colonialismo digital e seus impactos, o Brasil ainda não conseguiu garantir acesso básico à internet em suas escolas. Mesmo com R\$2,3 bilhões destinados pelo Ministério da Educação⁶ para conectividade no ensino básico (2023-2026), apenas 52,80% das escolas brasileiras (72.784) possuem conexão adequada. Com 65.130 unidades ainda desconectadas, o país se encontra numa posição vulnerável (Brasil, 2025).

Small language models como alternativa viável

Os LLMs (*Large Language Models*) ganharam destaque no campo da Inteligência Artificial devido à sua capacidade de gerar textos em linguagem natural. Contudo, apresentam limitações técnicas significativas na precisão das respostas, limitações operacionais relacionadas à alta demanda computacional por GPUs e RAM, impactos ambientais decorrentes da exploração de recursos naturais para resfriamento dos *datacenters* e da intensificação da extração de minérios, componentes fundamentais para o funcionamento dessas placas (ONU, 2024; Popov *et al.*, 2025).

Em meio ao caminho nebuloso da expansão das inteligências artificiais e suas operações de *datacenters* em países subdesenvolvidos operados pelas *bigtechs*, os SLMs (*Small Language Models*) emergem como solução tecnicamente viável para democratizar a representação científica do Sul Global em contextos de recursos limitados. Definidos como modelos com poucos milhões a alguns bilhões de parâmetros (geralmente <1-671 bilhões), oferecem vantagens substanciais sobre *Large Language Models* para aplicações educacionais no Sul Global.

⁶ Disponível em: <https://www.gov.br/casacivil/pt-br/assuntos/noticias/2025/maio/entenda-como-o-mec-monitora-a-conectividade-das-escolas>

Apesar das vantagens apresentadas sobre os SLMs, os estudos de Asai *et al.* (2024) mostraram menor eficiência e precisão das respostas dos modelos abertos de IAG, pois quanto menores os parâmetros, menor também será a precisão das respostas. Contudo, os mesmos autores afirmam que ao implementar a técnica RAG, as respostas e o raciocínio se tornam mais eficientes.

As vantagens técnicas incluem menor latência, menor consumo energético, capacidade de execução em *hardware* limitado e funcionamento *offline*, evitando assim a disponibilização dos dados dos usuários às grandes *bigtechs*. Nesse contexto, os SLMs podem ser treinados com dados de instituições locais de forma independente, sem necessidade de hospedagem em nuvens de grandes empresas, possibilitando uma alternativa viável para confrontar o colonialismo de dados e a sub-representação científica.

Estratégias decoloniais e caminhos alternativos

A literatura identifica três técnicas decoloniais para IA educacional: desenvolvimento técnico consciente do contexto, tutela reversa e solidariedade (Mohamed *et al.*, 2020). O desenvolvimento consciente do contexto requer que pesquisadores considerem onde e como sistemas serão utilizados, incluindo teste, políticas de apoio e organização contra usos impróprios.

Silveira *et al.* (2021) propõe estratégias concretas para resistência ao colonialismo de dados na educação. Em sua análise, o software livre e as tecnologias abertas são “brechas na estrutura do neoliberalismo e permitem a apropriação de tecnologias para a sua reconfiguração”, recebendo “influências das culturas e cosmovisões locais”.

Já as alternativas práticas incluem: pedagogia afrocentrada, integração de conhecimentos tradicionais de populações indígenas, quilombolas e ribeirinhas, e descentramento curricular para torná-los menos etnocêntricos e mais híbridos. Resistência mundana através de pequenos atos de recusa, resistência passiva inspirada na tradição radical negra, e organização local via *workshops*, artes e narrativas são exemplos que emergem como estratégia viável.

Considerações

O desafio da representação científica local em sistemas de IA constitui uma questão fundamental de justiça epistêmica e soberania tecnológica. A evidência apresentada, demonstra que a sub-representação sistemática do conhecimento do Sul Global não é meramente uma limitação técnica, mas resultado de estruturas coloniais que persistem na era digital. O *Small Language Models* emerge como alternativa tecnicamente viável e economicamente acessível para democratizar IA educacional, permitindo preservação e amplificação de saberes científicos locais.

A análise sobre colonialismo de dados fornece ferramentas conceituais para compreender como “entregamos nossos dados para o exterior” e como isso perpetua a dependência tecnológica. Sua proposta de soberania algorítmica oferece caminho concreto para desenvolvimento de sistemas de IA que sirvam aos interesses da inteligência coletiva local.

A transformação da educação tecnológica requer uma abordagem sistêmica que integre desenvolvimento de capacidades técnicas locais, formação docente contextualizada, epistemologias decoloniais e políticas de soberania digital. As diferenças metodológicas identificadas entre contextos anglo-saxões e não anglo-saxões não são deficiências a serem corrigidas, mas especificidades a serem valorizadas e incorporadas no desenvolvimento tecnológico.

A decolonização da IA educacional demanda não apenas mudanças técnicas, mas transformações estruturais nas relações de poder que governam desenvolvimento, implementação e governança tecnológica. Como argumenta Silveira *et al.* (2021), é urgente “iniciar um conjunto de pesquisas que tracem as redes de subordinação da colonialidade” e desenvolver tecnologias que “sirvam às localidades e aos interesses da inteligência coletiva”. O caminho para frente envolve solidariedade internacional, desenvolvimento de alternativas tecnológicas próprias e construção de epistemologias plurais que reconheçam múltiplas formas de conhecer e ser no mundo.

Referências

ASAI, Akari. *et al.* SELF-RAG: Learning to retrieve, generate, and critique through self-reflection. *arXiv*, 2023. Disponível em: <https://arxiv.org/abs/2310.11511>. Acesso em: 1 jul. 2025

BOAVENTURA, Julia da Silva. Colonialismo de dados: como opera a trincheira algorítmica na

guerra neoliberal. **O Social em Questão**, v. 1, n. 58, p.313-318, 2024. Disponível em: http://www.maxwell.vrac.pucrio.br/Busca_etds.php?strSecao=resultado&nrSeq=65398@1. Acesso em: 21 maio 2025.

CASA CIVIL. **Entenda como o MEC monitora a conectividade das escolas**. Casa Civil. 2025. Disponível em: <https://www.gov.br/casacivil/pt-br/assuntos/noticias/2025/maio/entenda-como-o-mec-monitora-a-conectividade-das-escolas> . Acesso em: 6 jul. 2025.

CRUZ, Leonardo Ribeiro da. *et al.* **Mapeamento da plataformização da educação pública superior: América Latina e África**. Observatório Educação Vigada, 2024. Disponível em: <https://zenodo.org/doi/10.5281/zenodo.11243189>. Acesso em: 1 jul. 2025.

FAUSTINO, Deivison; LIPPOLD, Walter. **Colonialismo digital: por uma crítica hackerfanoniana**. São Paulo: Boitempo, 2023. 218 p.

ISRAEL, Carolina Batista. *et al.* **Reconhecimento facial nas escolas públicas do Paraná**. Curitiba: Ibliblioteca de Ciência e Tecnologia, 2023. 142 p. Disponível em: https://observacuritiba.ufpr.br/wp-content/uploads-2023-12-rf_pr_2023-pdf-2/. Acesso em: 20 jun. 2025.

MANVI, Rohin. *et al.* Large Language Models are geographically biased. **arXiv**, 2024. Disponível em: <http://arxiv.org/abs/2402.02680>. Acesso em: 25 jun. 2025.

MONTE, Cidália Alves do. Tecnologias digitais na Educação: Vantagens, desafios e estratégias para uma integração eficiente no contexto brasileiro. **E-Acadêmica**, v. 6, n. 1, p. e0261600–e0261600, 2025. Disponível em: <<https://eacademica.org/eacademica/article/view/600>>. Acesso em: 7 jul. 2025

MOHAMED, Shakir; PNG, Marie-Therese; ISAAC, William. Decolonial AI: Decolonial Theory as sociotechnical foresight in Artificial Intelligence. **Philosophy & Technology**, v. 33, n. 4, p. 659–684, 2020.

POELL, Thomas; NIEBORG, David; DIJCK, José Van. Plataformização. **Fronteiras – Estudos Midiáticos**, v. 22, n. 1, p. 2–10, 2020. Disponível em: <http://revistas.unisinos.br/index.php/fronteiras/article/view/19838>. Acesso em: 1 jul. 2025.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence: a modern approach**. 5. ed. Reino Unido: Pearson Education, 2021

MARINI, R. M. Dialética da dependência. In: SADER, E. (Org.). *Dialética da dependência: uma antologia da obra de Ruy Mauro Marini*. Petrópolis: Vozes/CLACSO/Laboratório de Políticas Públicas, 2000. pp. 105-165. (Coleção A Outra Margem).

SANTOS, Boaventura de Sousa. **Justicia entre saberes: Epistemologías del Sur contra el epistemicidio**. Madrid: Morata, , 2017. 84 p.

SILVEIRA, Sérgio Amadeu da. Neocolonialismo ou imperialismo de dados? As novas veias abertas da América Latina?. In: BIRLE, Peter; WINDUS, Astrid (org.). **Conocimiento, poder y transformación digital en América Latina** Iberoamericana Vervuert, 2024. Disponível em: https://www.iberamericana-vervuert.es/capitulos/9783968694344_003.pdf. Acesso em: 6 jul. 2025.

UNITED NATIONS ENVIRONMENT PROGRAMME; INTERNATIONAL SCIENCE COUNCIL. **Navigating New Horizons: A global foresight report on planetary health and**

human wellbeing. United Nations Environment Programme, 2024. Disponível em: <https://wedocs.unep.org/20.500.11822/45890>. Acesso em: 4 jul. 2025.

VIEIRA, Livia Carolina; FERRARO, Danielle Soares E Silva Bicudo. Plataformização e precarização da experiência de aprendizagem na Educação Básica. **Caderno Pedagógico**, v. 22, n. 7, p. e16366, 2025

WANG, Guanzhi. *et al.* Voyager: An Open-Ended Embodied Agent with Large Language Models: **arXiv**, 2023. Disponível em: <https://arxiv.org/abs/2305.16291>. Acesso em: 7 jul. 2025.

YAN, Shi-Qi. *et al.* Corrective Retrieval Augmented Generation. **arXiv**, 2024. Disponível em: <https://arxiv.org/abs/2401.15884>. Acesso em: 6 jul. 2025.

ZUBOFF, Shoshana. **Capitalismo de vigilância**. Rio de Janeiro: Intrínseca, 2021. 824 p.