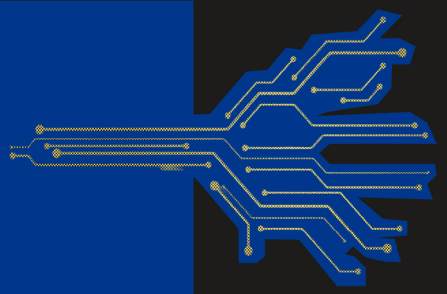




Instituto de
Pesquisa em
Direito & Tecnologia
do Recife















INOVAÇÃO PARA QUEM?

OS CUSTOS SOCIOAMBIENTAIS DA IA NO NORDESTE

Uma análise da infraestrutura e do ciclo de vida de modelos de inteligência artificial.

Inovação para quem? Os custos socioambientais da IA no Nordeste.
Uma análise da infraestrutura e do ciclo de vida de modelos de inteligência artificial.

Outubro de 2026

Instituto de Pesquisa em Direito e Tecnologia do Recife – IP.rec

Este relatório integra o projeto sinérgico dos "Desafios Emergentes da Inteligência Artificial", desenvolvido pelo IP.rec, com apoio financeiro da Fundação Heinrich Böll - Brasil, do Fundo CASA e da Luminate.

Pesquisa e redação
André Lucas Fernandes
Fabiana Teixeira
Lauro Accioly

Coordenação e revisão
André Lucas Fernandes
Anicely Santos

Projeto gráfico
Estúdio PUYA!

Este conteúdo está licenciado sob a Creative Commons Atribuição-Não Comercial 4.0 Internacional (CC BY-NC 4.0), permitindo o uso, compartilhamento e adaptação apenas para fins não comerciais, com atribuição à autoria original.

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Fernandes, André Lucas
Inovação para quem? : os custos socioambientais da IA no nordeste [livro eletrônico] : uma análise da infraestrutura e do ciclo de vida de modelos de inteligência artificial : Instituto de Pesquisa em Direito e Tecnologia do Recife / André Lucas Fernandes, Fabiana Teixeira, Lauro Accioly ; coordenação André Lucas Fernandes, Anicely Santos.
-- Recife, PE : IP.rec, 2026.
PDF

ISBN 978-65-982630-6-5

1. Inovação tecnológica 2. Inteligência artificial 3. Sustentabilidade 4. Tecnologia I. Teixeira, Fabiana. II. Accioly, Lauro. III. Santos, Anicely. IV. Título.

26-408631.0 CDD-006.3

Índices para catálogo sistemático:

1. Inteligência artificial 006.3

Livia Dias Vaz - Bibliotecária - CRB-8/9638

Introdução



O debate sobre os impactos ambientais associados à infraestrutura necessária ao desenvolvimento e operação de sistemas de Inteligência Artificial (IA) tem ganhado crescente relevância. Embora a literatura venha avançando na identificação desses impactos, ainda existem lacunas relacionadas à sua dimensão socioambiental quando considerados diferentes tipos de modelos de IA e os contextos territoriais. Nesse sentido, este *brief(ing) paper* busca contribuir para o aprofundamento desse debate a partir de um recorte que articula o ciclo de vida dos modelos de IA, com atenção às diferenças entre grandes modelos (*large models*) e pequenos modelos (*small models*) e as particularidades socioambientais da região Nordeste no Brasil.

A pesquisa tem como objetivo compreender, por meio de uma revisão de literatura, os principais impactos socioambientais que podem advir da implementação de massiva infraestrutura de IA no Nordeste brasileiro, considerando diferentes etapas do ciclo de vida dos modelos, desde seu desenvolvimento e treinamento até sua implementação e uso. Para isso, foram adotados procedimentos sistemáticos de busca e seleção de trabalhos científicos nas plataformas *Google Scholar* e *Web of Science*. Os processos utilizados seguiram as orientações de revisão de literatura de Flor *et al* (2021), focando na elaboração de uma síntese em diferentes tópicos, a fim de gerar uma compreensão mais ampla sobre o conhecimento estudado.[1]

Em uma primeira etapa, os estudos identificados foram organizados a partir de três escopos de recortes para busca e seleção. O primeiro focou no caráter mais amplo do tema e visou o macro tema "os impactos socioambientais da Inteligência Artificial". O segundo focou no caráter mais específicos atrelados aos diferentes tipos e dimensões dos modelos de IA, com destaque para *large models* e *small models*. Por fim, o terceiro foi voltado para os impactos socioambientais da infraestrutura de IA associados a esses modelos especificamente no contexto da região Nordeste no Brasil.

Para cada escopo, foram definidos conjuntos específicos de termos de busca, utilizados para identificar e filtrar os trabalhos nas duas plataformas selecionadas,

---

1 FLOR, Tainá de Oliveira et al. **Revisões de literatura como métodos de pesquisa: aproximações e divergências. Plataforma Espaço Digital. Anais do VI CONAPESC.** Campina Grande: Realize Editora. Publicado em, v. 24, 2021.

o recorte temporal em toda a pesquisa foi o mesmo: 4 anos (2022-2025)[2]. A seguir, apresentamos os termos empregados em cada escopo, bem como os resultados quantitativos obtidos a partir das buscas. Essa sistematização permite visualizar a distribuição da produção científica encontrada e, posteriormente, discutir as lacunas identificadas na literatura em relação ao recorte proposto.

O trabalho de pesquisa aqui compreendido está inserido na modalidade de pesquisa-ação, com fins de subsidiar discussões para construção de políticas públicas na temática. É por este motivo que há uma hipótese subjacente a ser discutida: pequenos modelos seriam alternativas viáveis para escapar de um processo erosivo de construção, treinamento e uso de grandes modelos de IA, especialmente a generativa[3].

Escopo 1: Os impactos socioambientais da IA

A seleção bibliográfica foi realizada em dois idiomas, português e inglês, a partir de diferentes combinações de termos relacionadas à infraestrutura de inteligência artificial e seus impactos ambientais e socioambientais. Nas buscas em inglês, foram utilizadas inicialmente duas combinações de termos. A primeira, "AI infrastructure" + "environmental impact", resultou em 1.680 registros no Google Scholar e 363 na Web of Science. A segunda combinação, "AI model" + "environmental footprint", identificou 1.920 registros no Google Scholar e 256 na Web of Science.

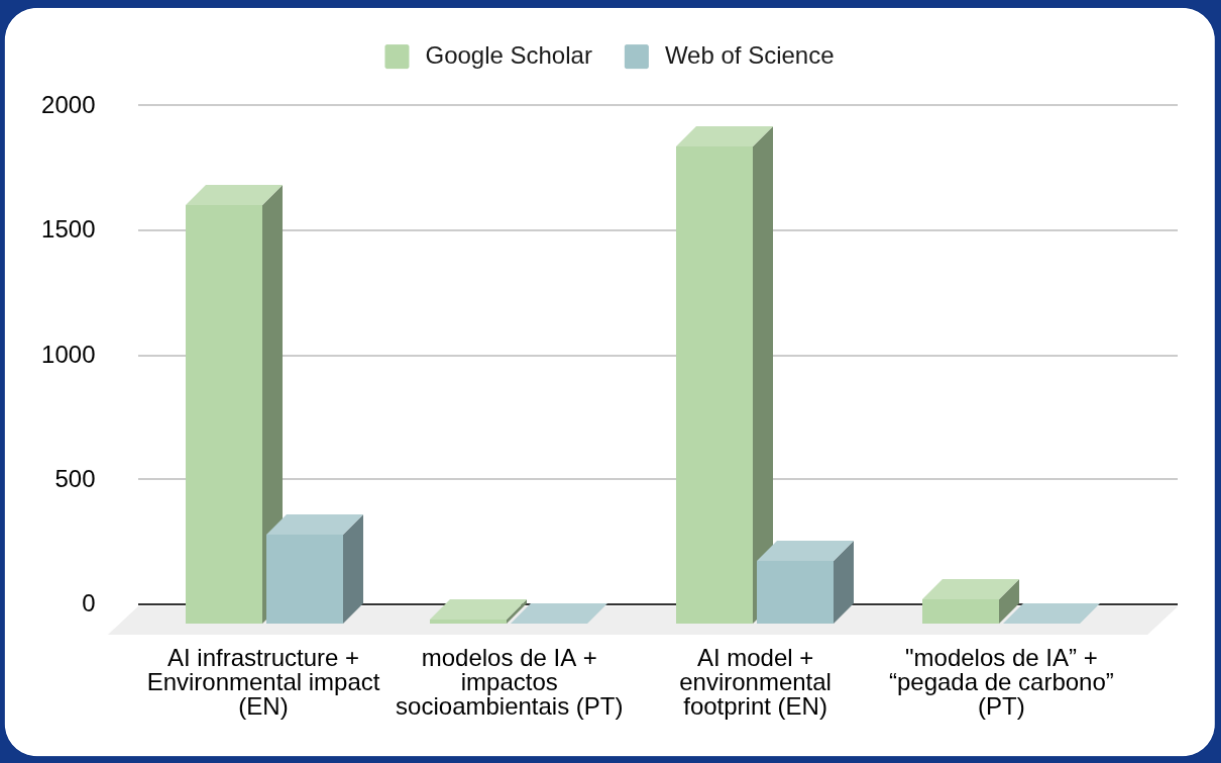
Nas buscas em português, a combinação "modelos de IA" + "impactos socioambientais" resultou em 21 produções científicas no Google Scholar, não sendo identificados registros na Web of Science. Em uma busca mais específica, voltada à relação entre inteligência artificial, emissões de carbono e o contexto regional, foram utilizados os termos "modelos de IA" + "pegada de carbono" + "Nordeste brasileiro", que retornaram 99 publicações no Google Scholar e nenhum registro na Web of Science.

---

2 Todas as filtragens seguiram o recorte temporal de quatro anos (2022–2025). O ano vigente (2026) não foi incluído, considerando que novas pesquisas e publicações ainda poderão ser disponibilizadas ao longo do ano. A exclusão busca garantir maior consistência e comparabilidade do conjunto de dados, evitando incompatibilidades decorrentes da incorporação posterior de novos estudos.

3 WANG, Libing; WANG, Tianchong. **Small language models (SLMs): A cheaper, greener route into AI.** UNESCO, 18 mar. 2024. Atualizado em: 11 nov. 2025. Disponível em: https://www.unesco.org/en/articles/small-language-models-slms-cheaper-greener-route-ai. Acesso em: 13 out. 2026.

Gráfico 1. Resultados das buscas no Escopo 1 (combinação de termos e base de dados)



Fonte: elaborado pelos autores

Os resultados evidenciam diferenças expressivas no volume de publicações recuperadas de acordo com o idioma, a base de dados e o grau de especificidade dos termos empregados, contribuindo para a delimitação do escopo da revisão e para a identificação da produção científica relacionada ao tema.

Escopo 2. Impactos Ambientais da IA com foco nos Modelos

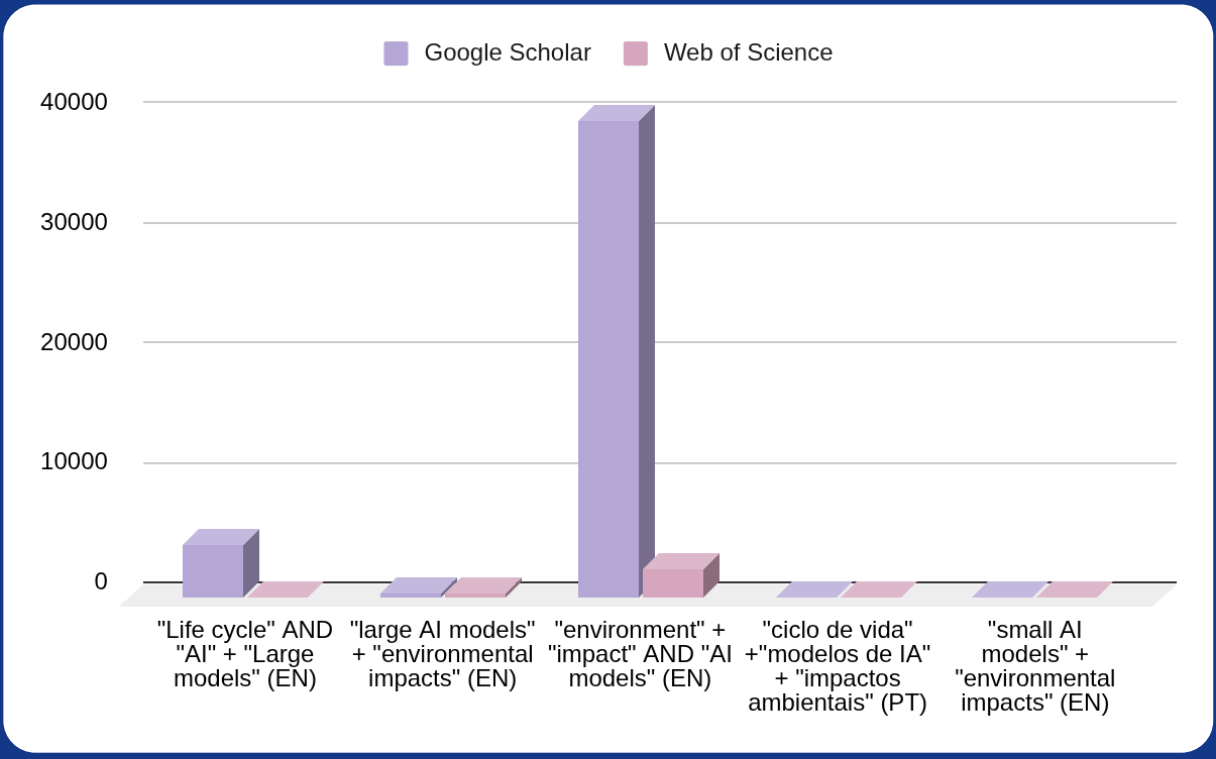
No segundo escopo, denominado "Impactos Ambientais da IA com foco nos Modelos", as buscas foram direcionadas à identificação da produção científica relacionada aos impactos ambientais associados aos modelos de inteligência artificial, considerando diferentes níveis de especificidade dos termos de busca. Foram utilizadas combinações de palavras-chave em inglês e português, com consultas realizadas nas bases Google Scholar e Web of Science.

Entre as buscas em inglês, a combinação "Life cycle" + "AI" + "Large models" resultou em 4.370 registros no Google Scholar e 143 na Web of Science. Já a busca pelos termos "large AI models" + "environmental impacts" identificou 363 registros no Google Scholar e 369 na Web of Science. Uma combinação mais

ampla, utilizando os termos “environment” + “impact” + “AI models”, apresentou 39.800 registros no Google Scholar e 2.363 na Web of Science, evidenciando um aumento expressivo no número de resultados quando empregados termos de busca mais abrangentes.

Em português, a combinação “Ciclo de vida” + “Modelos de IA” + “Impactos Ambientais” resultou em 94 registros no Google Scholar e nenhum registro na Web of Science. Também foi realizada uma busca específica pelos termos “Small AI models” + “Environmental Impacts”, em inglês, que identificou 4 registros no Google Scholar e nenhum na Web of Science.

Gráfico 2. Resultados das buscas no Escopo 2 (combinação de termos e base de dados)



Fonte: elaborado pelos autores

Os resultados desse escopo demonstram que o volume de publicações recuperadas varia significativamente conforme a abrangência e a especificidade das combinações de termos utilizadas. As buscas mais amplas apresentaram maior número de registros, enquanto aquelas direcionadas a aspectos específicos dos modelos de IA, como ciclo de vida e impactos ambientais, resultaram em um conjunto mais delimitado de publicações. Esses resultados contribuíram para o refinamento do escopo e para a identificação das abordagens mais recorrentes na literatura sobre os impactos ambientais dos modelos de inteligência artificial.

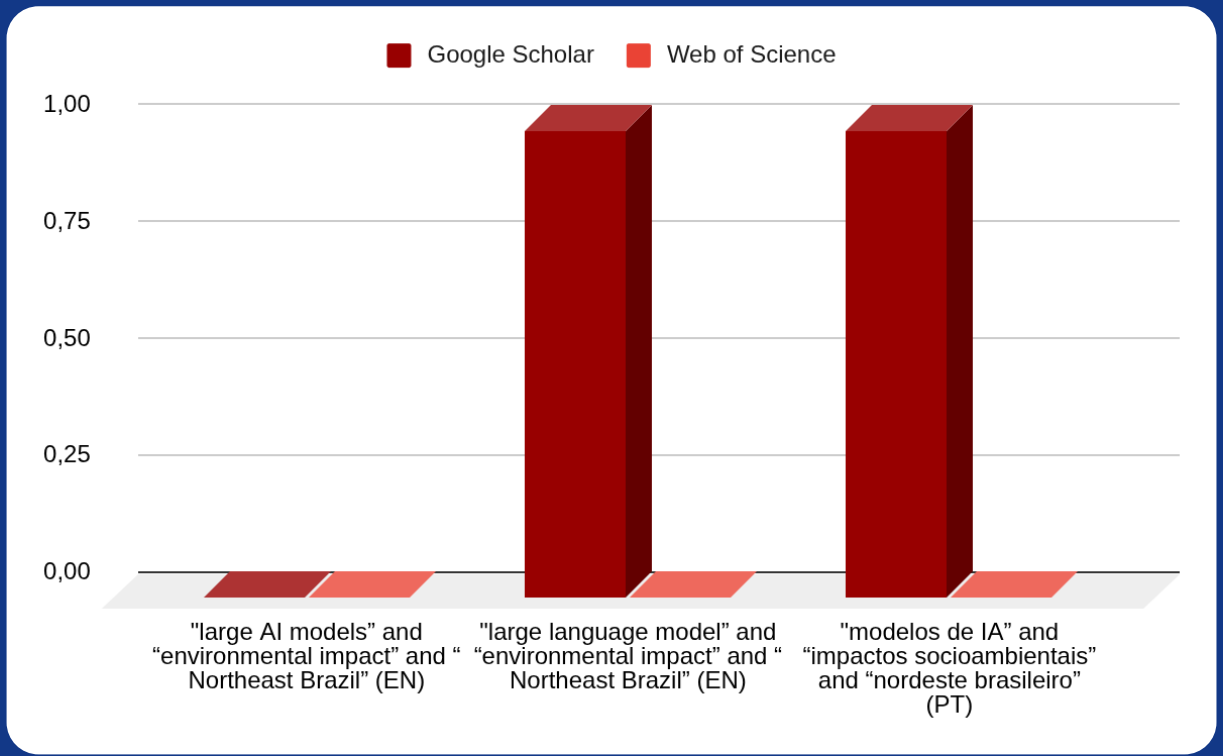
Escopo 3. Impactos ambientais da IA com foco em Modelos no Nordeste brasileiro

No terceiro escopo, a busca foi direcionada à identificação de produções científicas que relacionassem os impactos ambientais e socioambientais dos modelos de inteligência artificial ao contexto específico do Nordeste brasileiro. Para isso, foram realizadas buscas em inglês e português nas bases Google Scholar e Web of Science, incorporando o recorte geográfico às combinações de termos.

Na busca em inglês, a combinação "large AI models" AND "environmental impact" AND "Northeast Brazil" não apresentou resultados em nenhuma das bases consultadas, com 0 registros no Google Scholar e 0 na Web of Science. A combinação "large language model" AND "environmental impact" AND "Northeast Brazil" identificou apenas 1 registro no Google Scholar, não sendo encontrado nenhum registro na Web of Science.

Em português, a combinação "modelos de IA" AND "impactos socioambientais" AND "Nordeste brasileiro" também apresentou apenas 1 registro no Google Scholar e nenhum registro na Web of Science.

Gráfico 3. Resultados das buscas no Escopo 3 (combinação de termos e base de dados)



Fonte: elaborado pelos autores

Os resultados evidenciam um número bastante reduzido de publicações que articulam simultaneamente os modelos de inteligência artificial, seus impactos ambientais ou socioambientais e o recorte específico do Nordeste brasileiro. A ausência ou baixa ocorrência de registros nas bases consultadas indica a necessidade de considerar o caráter ainda restrito dessa produção científica no contexto regional, além de reforçar a importância de estratégias de busca complementares para a identificação de estudos relacionados ao tema.

Escopo 4. Impactos ambientais da IA com foco em Modelos no Brasil

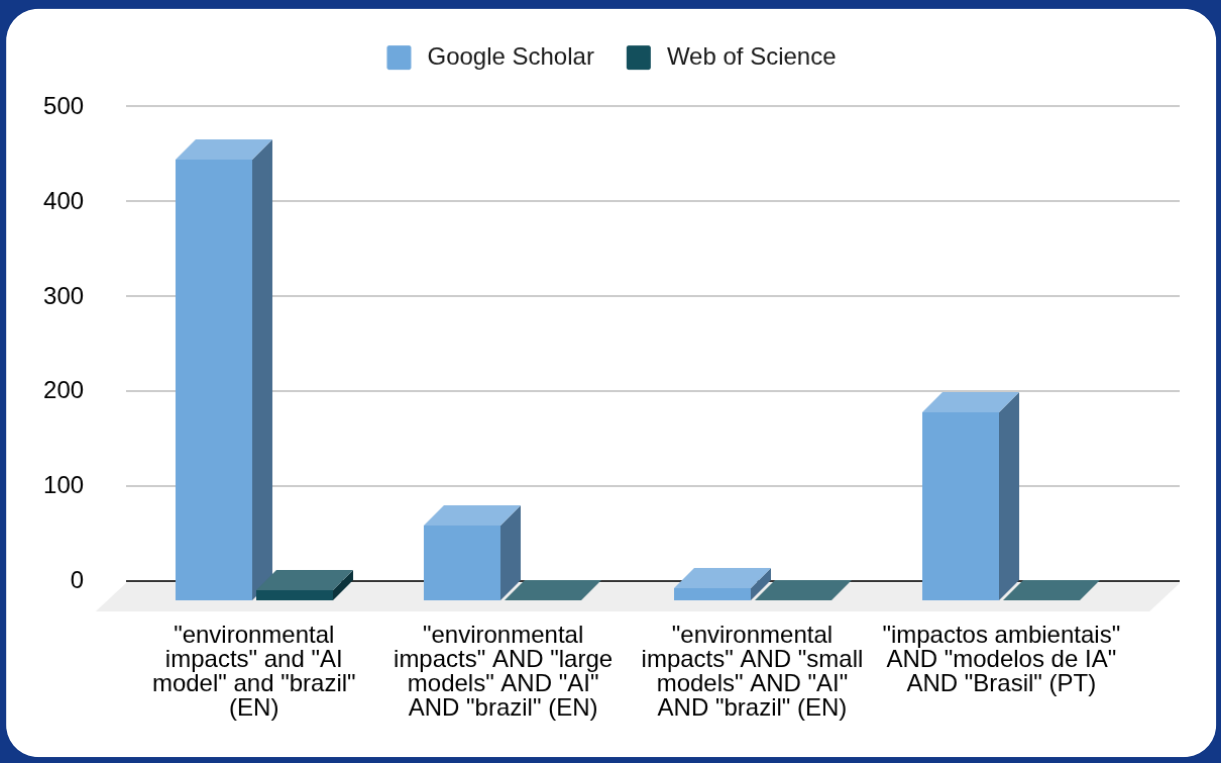
A partir dos resultados obtidos nas etapas anteriores, identificou-se uma limitação importante no que diz respeito à representação geográfica da produção científica sobre os impactos ambientais da Inteligência Artificial. Apesar do interesse inicial em investigar a temática com ênfase na região Nordeste, a quantidade reduzida de estudos recuperados com essa delimitação indicou que a produção científica disponível apresenta pouca atenção explícita a esse recorte regional. Diante dessa lacuna, foi realizada uma etapa adicional de busca, ampliando o escopo geográfico para todo o território brasileiro, com o intuito de verificar se a baixa ocorrência de estudos estava associada à própria escassez de pesquisas sobre o tema no contexto nacional ou se decorria, especificamente, da delimitação ao Nordeste.

Nessa etapa, foram realizadas buscas no Google Scholar e na Web of Science, utilizando termos em português e inglês relacionados aos impactos ambientais, aos modelos de IA e ao Brasil. Foram empregadas estratégias envolvendo o termo geral “AI model”, bem como diferenciações entre “large models” e “small models”. No Google Scholar, a combinação “environmental impacts” + “AI model” + “Brazil” retornou 466 resultados, enquanto as buscas envolvendo especificamente “large models” e “small models” retornaram 80 e 14 resultados, respectivamente. A busca em português, “impactos ambientais” + “modelos de IA” + “Brasil”, apresentou 200 resultados. Na Web of Science, por sua vez, foram recuperados apenas 12 resultados para a primeira estratégia, 1 para a busca envolvendo “large models” e nenhum resultado para “small models” ou para a estratégia em português.

A ampliação do recorte permitiu evidenciar um aspecto relevante da própria distribuição da literatura: quando a busca é direcionada especificamente ao Nordeste, a quantidade de trabalhos recuperados é ainda mais restrita, enquanto a adoção de um recorte nacional amplia consideravelmente o número de resultados, sobretudo no Google Scholar. Esse contraste sugere que a discussão sobre os impactos ambientais associados aos modelos de IA no Brasil ainda não apresenta uma atenção regional equilibrada, com o Nordeste aparecendo de maneira pouco

expressiva nas estratégias de busca realizadas. Assim, a escassez de resultados não deve ser interpretada necessariamente como ausência de impactos ou de relevância do fenômeno na região, mas como um indicativo da limitada produção científica explicitamente localizada nesse território, o que torna ainda mais sensível a fronteira de expansão de centros de dados nesta região, associados à matriz energética renovável já instalada.

Gráfico 4. Resultados das buscas no Escopo 4 (combinação de termos e base de dados)



Fonte: elaborado pelos autores

Nesse sentido, essa etapa extra foi incorporada como uma estratégia exploratória para tornar visível essa lacuna e avaliar o quanto a ampliação da escala geográfica modifica a recuperação da literatura. O procedimento também reforça a necessidade de considerar possíveis assimetrias regionais na produção e na localização dos estudos sobre os impactos ambientais da IA, uma vez que uma literatura predominantemente orientada para o contexto brasileiro em escala nacional pode não representar adequadamente especificidades e demandas de regiões como o Nordeste.

Em consideração de possíveis não indexação de publicações acadêmicas universitárias de Pernambuco[4] nessas plataformas, ainda tentamos uma busca nos repositórios universitários para uma dupla checagem, todavia, não foi possível encontrar nenhum resultado seguindo o recorte temporal e os termos usados na busca dos escopos 3 e 4.



4 Os repositórios universitários selecionados foram referentes as seguintes universidades: UPE (Estadual), UFRPE (Federal Rural) e UFPE (Federal).

2.
Os impactos da infraestrutura da Inteligência Artificial



A partir do levantamento bibliográfico feito no escopo 1, foi analisado que os argumentos de busca mais abrangentes como "AI infrastructure + environmental impacts" e "AI model + environmental footprint" forneceram maior quantidade de artigos científicos sobre a referida temática no Google Scholar e Web of Science, escritos majoritariamente em inglês. Nesse sentido, ao utilizar termos em português e centralizados no nordeste brasileiro, como "modelos de IA + impactos ambientais" e "modelos de IA + pegada de carbono", o número de resultados encontrados reduziu significativamente, evidenciando uma carência na produção científica sobre os impactos socioambientais atrelados à infraestrutura da IA na região.

Neste momento, foram selecionados os quatro artigos com mais citações nas fontes de busca utilizadas, filtrados pela sua relevância acadêmica e pertinência temática. A tabela a seguir evidencia os artigos selecionados para esta etapa, encontrados através dos argumentos de busca do escopo 1 mencionados acima.

PLATAFORMA DE BUSCA	TÍTULO	CITAÇÕES
GOOGLE SCHOLAR	MEASURING THE ENVIRONMENTAL IMPACT OF DELIVERING AI AT GOOGLE SCALE	95
GOOGLE SCHOLAR	SUSTAINABLE AI: ENVIRONMENTAL IMPLICATIONS, CHALLENGES AND OPPORTUNITIES	1370
GOOGLE SCHOLAR	OUR NEW ARTIFICIAL INTELLIGENCE INFRASTRUCTURE: BECOMING LOCKED INTO AN UNSUSTAINABLE FUTURE	139
WEB OF SCIENCE	TOWARDS SUSTAINABLE AI: A COMPREHENSIVE FRAMEWORK FOR GREEN AI	60

Para analisar o contexto do nordeste do Brasil, procuramos compreender os impactos socioambientais atrelados à infraestrutura da IA. Robbins e Van Wynsberghe[5] buscam analisar a IA como uma infraestrutura da qual muitos serviços dependem, e não somente como uma ferramenta de expansão das infraestruturas já

5 ROBBINS, Scott; VAN WYNSBERGHE, Aimee. Our new artificial intelligence infrastructure: becoming locked into an unsustainable future. **Sustainability**, v. 14, n. 8, p. 4829, 2022.

existentes. Ainda que se utilize a palavra "nuvem" para se referir à IA, desatrelando sua materialidade de uma estrutura física, há uma infraestrutura física por trás de sua produção, que é dependente de inúmeras consequências ambientais opacas. Por isso, os autores abordam o campo da IA sustentável como uma forma de discutir a justiça ambiental, visando promover mudanças no ciclo de vida da IA.

Nesta seção, mencionamos os seguintes custos ambientais: (I) consumo de energias não-renováveis; (II) emissões de carbono; (III) consumo de recursos hídricos; (IV) descarte de resíduos eletrônicos; (V) e exploração de recursos minerais preciosos, considerando a infraestrutura de prestação de serviços de IA. Para Robbins e van Wynsberghe (2022), a IA não somente contribui para a emissão direta de carbono, mas sua infraestrutura requer a extração de minerais preciosos, extraídos mediante condições precárias de trabalho. No mesmo sentido, a necessidade de exploração de elementos de terras raras também gera impactos ambientais significativos.

A criação de modelos de IA de grande escala emite aproximadamente 626.000 libras de CO_2, exigindo uma elevada quantidade de poder computacional. Ao mesmo tempo, isso requer mais hardware, o que, por seu turno, gera processos de consumo em níveis elevados de energia. Por conseguinte, o descarte e a reciclagem do hardware, feito de maneira constante, geram lixo eletrônico (e-lixo)[6], frequentemente descartado em países subdesenvolvidos, expondo as comunidades locais à resíduos tóxicos em contato direto com a água e o solo[7]. Dessa forma, Robbins e Van Wynsberghe (2022) explicam que a sociedade tem criado uma relação de dependência com a economia da IA, aprisionando-se em uma infraestrutura que possui impactos ambientais significativos.

Em estudo de caso do *Gemini*, Elsworth *et al.*[8] buscaram analisar o consumo de energia, as emissões de carbono e o consumo de água das infraestruturas associadas aos serviços de IA de larga escala, bem como os esforços do Google para aquisição de energia limpa e mitigação de impactos ambientais. Em 2025, foi constatado que um prompt de texto no Gemini consome aproximadamente 0,24 Wh de energia, que provém da operação do acelerador de IA (58% do total), do consumo da CPU e DRAM do Host (25%), de energias consumidas por máquinas provisionadas em estado ocioso (10%) e a sobrecarga do data center (8%). Para estes autores, ainda

---

6 ZHUK, Alesia. Artificial intelligence impact on the environment: hidden ecological costs and ethical-legal issues. **Journal of Digital Technologies and Law**, v. 1 n. 4, p. 932-954, dez. (2023). https://doi.org/10.21202/jdtl.2023.40

7 BADAWI, Ibrahim. E-waste, artificial intelligence, and the law: toward a circular and sustainable digital economy. **SSRN**, set. (2025). https://doi.org/10.2139/ssrn.6012214.

8 Elsworth, C., Huang, K., Patterson, D., Schneider, I., Sedivy, R., Goodman, S., Townsend, B., Ranganathan, P., Dean, J., Vahdat, A., et al. (2025). **Measuring the environmental impact of delivering AI at Google Scale**.

que o consumo de energia seja um fator importante, os impactos ambientais são gerados principalmente pelas emissões e consumo de água.

Aqui, o texto analisado propõe a redução da intensidade das emissões e a limitação do uso de água em bacias hidrográficas de alto estresse hídrico como mecanismos de redução significativa do impacto da infraestrutura da IA ao meio ambiente. Para eficiência de emissões, deve-se criar uma combinação de esforços para arquitetura de modelos mais inteligentes, algoritmos e quantização eficientes, inferência e fornecimento otimizados, bem como a criação de data centers ultraeficientes e aquisição de energia limpa.

Para *Tabbakh et al.*[9], os data centers que abrigam a infraestrutura da IA são grandes consumidores de água e emissores de GEE (gases de efeito estufa). Portanto, o impacto ambiental da IA requer a adoção de práticas sustentáveis de pesquisa e desenvolvimento, ampliando sua eficiência energética e promovendo o uso de recursos ecologicamente corretos, a partir do desenvolvimento de algoritmos energicamente eficientes, da otimização de arquiteturas de modelos e do aproveitamento de fontes de energia renováveis para operações de data centers. Assim, os autores citam a *Green AI* (IA Verde), que refere-se à implementação de tecnologias de IA alinhadas com os Objetivos de Desenvolvimento Sustentável (ODS) e buscam reduzir os impactos ambientais, com a diminuição do consumo de energia e emissões de carbono. Para tanto, seus critérios específicos são: (I) eficiência energética e uma pegada de carbono reduzida; (II) a otimização de algoritmos de IA; (III) a utilização de hardware energeticamente eficiente; (IV) e a utilização de fontes de energia renováveis para alimentar as operações de IA.

Segundo Tabbakh *et al.*, o treinamento de modelos de IA em larga escala exige uma capacidade computacional elevada, resultando no consumo massivo de energia. Ademais, os data centers que abrigam a infraestrutura da IA também são grandes consumidores de eletricidade. Para mitigar os impactos da IA nas mudanças climáticas é necessário reduzir o consumo de energia e a pegada de carbono, e práticas sustentáveis de IA podem contribuir diretamente para essa redução. Assim, as técnicas de **poda**, que reduz o tamanho do modelo e acelera a computação, de **aumento de dados**, que diminui o consumo de energia, e de **quantização**, que reduz a precisão numérica dos parâmetros, levando a diminuição do tamanho do modelo, à aceleração da computação e à eficiência energética, são técnicas mencionadas para otimização dos modelos de IA.[10] Segundo os autores, a adoção de fontes de energia

---

9 Tabbakh, A., Al Amin, L., Islam, M. et al. Towards sustainable AI: a comprehensive framework for Green AI. **Discovery Sustainability 5**, 408 (2024). https://doi.org/10.1007/s43621-024-00641-4

10 BOLÓN-CANEDO, Verónica et al. A review of green artificial intelligence: towards a more sustainable future. **Neurocomputing**, v. 599, art. 128096, set. (2024). https://doi.org/10.1016/j.neucom.2024.128096

renováveis (como solar e eólica) nas operações de data centers também contribui significativamente para a redução nas emissões de carbono.

O estudo de Wu *et al.*[11] busca explorar os impactos ambientais da IA por uma perspectiva holística, levando em consideração o ciclo de vida do modelo e o hardware do sistema, incluindo a fabricação e o uso operacional. Para os autores, a otimização da eficiência em todos os eixos (algoritmos, plataformas, infraestrutura e hardware) pode reduzir de maneira significativa a pegada de carbono operacional da infraestrutura da IA.

Além disso, o estudo defende uma mentalidade sustentável para a IA, em que a comunidade científica deve buscar compreender e mitigar seus custos ambientais, utilizando-se de métricas de sustentabilidade ambiental como critérios de avaliação de pesquisas em aprendizado de máquina (*machine learning*), exigindo, por exemplo, declarações de impacto das emissões na infraestrutura da IA. Nesse sentido, utilizar metodologias de contabilização de carbono e telemetria destaca-se como uma das propostas para quantificar a relevância do progresso no desenvolvimento de tecnologias de IA com critérios ambientalmente responsáveis. Assim, métricas de sustentabilidade, como consumo de energia e pegada de carbono, juntamente com medidas de qualidade do modelo e desempenho do sistema são essenciais para análise de inovação e qualidade de soluções de aprendizado de máquina.

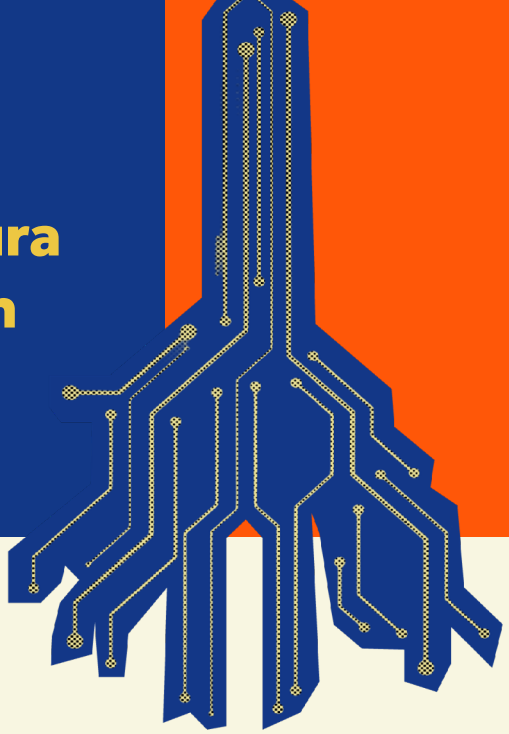
Algumas das soluções apresentadas por Wu *et al.* para mitigação dos impactos ambientais, baseadas no caso Meta, analisado no texto, são as seguintes: (I) (co) design holístico de hardware e software, que ilustrou uma redução de mais de 800 vezes na pegada de carbono operacional; (II) otimização de algoritmos e arquitetura de modelos por quantização e compressão de memória; (III) eficiência de dados e amostragem inteligente, aprimorando a análise competitiva para algoritmos de aprendizado de máquina, diminuindo a pegada ambiental; (IV) otimização do *pipeline* e infraestrutura de *data centers*; (V) e ampliação da vida útil do hardware, para combater o desgaste do hardware antigo e descarte precoce. Somado a isso, como exposto anteriormente, Wu *et al.* defendem uma mentalidade de sustentabilidade para IA com otimização além do convencional, com eficiência em toda a pilha de software e hardware.

---

11 WU, Carole-Jean et al. Sustainable AI: Environmental Implications, Challenges and Opportunities. Proceedings of the **5th MLSys Conference, Santa Clara, CA, USA**, 2022.



3. Os impactos da infraestrutura da Inteligência Artificial com base nos modelos (*large and small models*)



Em razão do expressivo volume de produções científicas na etapa inicial da revisão, procedeu-se a uma nova etapa de filtragem do *corpus*, com o objetivo de delimitar os estudos mais pertinentes à composição deste *paper*. Em consonância com as etapas de leitura propostas por Gil (2008)[12], adotamos uma abordagem exploratória, priorizando, nesta seção, os quatro artigos com maior número de citações entre os resultados identificados na filtragem feita no escopo 2. A utilização deste segundo critério foi pensado para estabelecer um recorte orientado pela relevância acadêmica dos trabalhos, contribuindo para a organização e síntese das evidências acerca dos impactos associados à infraestrutura necessária ao desenvolvimento e à operação de modelos de inteligência artificial, tanto de grande escala (*large models*) quanto de menor escala (*small models*).

Cabe destacar que, em alguns casos, os registros de citações não estavam integralmente disponíveis na Web of Science, o que impôs limitações à aplicação uniforme desse critério de seleção. Nesses casos, foram considerados os dados de citação disponíveis nas bases consultadas, preservando-se, tanto quanto possível, a consistência do procedimento de seleção. A partir desse recorte, a seção concentra-se nos principais impactos identificados na literatura, considerando as especificidades de infraestruturas associadas a diferentes escalas de modelos de IA.[13]

A seguir, projetamos uma tabela para tornar mais visível os artigos que foram selecionados, com base no termo **“Large AI models” + “Environmental impacts”**:

---

12 GIL, A. C. Métodos e técnicas de pesquisa social. 6 ed. São Paulo: Atlas, 2008.

13 Em razão do formato deste documento ser mais sucinto, decidimos abordar apenas os textos mais citados dos termos que são diretamente focados em impactos ambientais com base na especificação dos modelos: large and small models. Desta maneira, só serão utilizados os 4 textos mais citados dos dois termos filtrados.

PLATAFORMA DE BUSCA	TÍTULO	CITAÇÕES
WEB OF SCIENCE	PREVENTING THE IMMENSE INCREASE IN THE LIFE-CYCLE ENERGY AND CARBON FOOTPRINTS OF LLM-POWERED INTELLIGENT CHATBOTS	77
WEB OF SCIENCE	ENERGY-AWARE MACHINE LEARNING MODELS—A REVIEW OF RECENT TECHNIQUES AND PERSPECTIVES	79
GOOGLE SCHOLAR	OUR NEW ARTIFICIAL INTELLIGENCE INFRASTRUCTURE: BECOMING LOCKED INTO AN UNSUSTAINABLE FUTURE	139
GOOGLE SCHOLAR	ARTIFICIAL INTELLIGENCE COLONIALISM: ENVIRONMENTAL DAMAGE, LABOR EXPLOITATION, AND HUMAN RIGHTS CRISES IN THE GLOBAL SOUTH	89

A conjuntura atual de expansão dos modelos de Inteligência Artificial de larga escala (nomeados como large scale models) resulta em implicações ambientais devido à crescente demanda por elevado poder computacional, contribuindo para uma pegada de carbono significativa. Por exemplo, segundo Rózycki, Solarska e Waligóra[14], o treinamento de um modelo BERT pode gerar emissões de CO_2 equivalentes às de 5 carros ao longo de toda sua vida útil. Todavia, o impacto ambiental vai além das emissões operacionais, já que inclui emissões incorporadas (aquelas associadas à fabricação, transporte, descarte de hardware necessário à operação). A expansão de modelos como GPT-3 revela essas questões. Outro destaque é referente a complexidade do modelo favorecer uma ampliação da pegada ambiental da IA.

Neste aspecto, conforme apontado na pesquisa de Jiang *et al.*[15], o debate sobre os impactos ambientais atrelados aos large language models (LLMs) precisa ser expandido para compreender uma abordagem de ciclo de vida, indo além de analisar emissões associadas apenas ao treinamento dos modelos. A pesquisa estimou que o treinamento do GPT-3 consome aproximadamente 1.287MWh de eletricidade e gerou 552 toneladas de CO_2 e, enquanto o GPT-4 estaria em cerca de 7.200 MWh com emissão de 3.088 toneladas de CO_2 - um aumento de mais de 450% (quatrocentos e cinquenta por cento). Os desafios adicionais de resfriamento dos servidores foram por causa da alta necessidade computacional de operação desses modelos, cujas estimativas de consumo hídrico seriam de 5.4 milhões de litros de água referente ao GPT-3. Outrossim, seria o consumo mensal de eletricidade capaz de alcançar 23.364 MWh.

---

14 RÓŻYCKI, Rafał; SOLARSKA, Dorota Agnieszka; WALIGÓRA, Grzegorz. Energy-aware machine learning models—a review of recent techniques and perspectives. **Energies**, v. 18, n. 11, p. 2810, 2025

15 JIANG, Peng et al. Preventing the immense increase in the life-cycle energy and carbon footprints of LLM-powered intelligent chatbots. **Engineering**, v. 40, p. 202-210, 2024.

Por outro lado, Robbins e Van Wynsberghe[16] ampliam a discussão com a introdução do conceito *carbon lock in* (aprisionamento em carbono), destacando como investimentos de longo prazo em centros de dados, equipamentos computacionais e outras infraestruturas necessárias ao ciclo de vida da IA podem reforçar a dependência de sistemas intensivos em energia e consumo hídrico.

Contemplando os pontos já apontados anteriormente, Regilme[17] vai adicionar ao debate que grandes empresas de tecnologia acabam privilegiando instalar data centers em regiões da África e Sudeste Asiático devido às possibilidades de acesso a eletricidade mais barata e menor rigor na fiscalização ambiental. A dinâmica observada permite constatar assimetrias nos processos de geração de impactos, cujos benefícios econômicos da IA se concentram nos países do Norte global enquanto os do Sul global recebem maiores pesos de custos socioambientais.

No caso da filtragem de busca pelos termos **"Small AI models" + "Environmental impacts"**, não foi possível selecionar as quatro produções científicas mais citadas devido o baixo número de materiais achados, indicando a necessidade de mais pesquisa científicas que explorem as implicações do uso de pequenos modelos de IA e suas possíveis correlações com os impactos ambientais.

Deste modo, com base nos materiais encontrados, um ponto destacável da distinção entre os modelos e suas relações com os impactos ambientais é que a adequação do hardware às demandas computacionais de cada aplicação importa. De acordo com Fayza *et al.*[18], usos de aceleradores baseados em CMOS, com área inferior a 10 mm² e consumo de energia na ordem de miliwatts podem atender aplicações com eficiência energética, ao contrário de aceleradores eletrofotônicos que ocupam mais de 100 mm² e consomem dezenas de watts.

Segundo Vuruma *et al.*[19], há técnicas de reeducação dos recursos computacionais atreladas aos pequenos modelos de linguagem (*small language models*) que permitem reduzir os possíveis impactos ambientais: 1) ***model pruning***: remover componentes redundantes ou pouco relevantes ao funcionamento do modelo; 2) ***quantization***: representação dos parâmetros com menor precisão, reduzindo o tamanho do modelo e o uso de memória; 3) ***knowledge distillation:*** transferir conhecimento de

16 ROBBINS, Scott; VAN WYNSBERGHE, Aimee. Our new artificial intelligence infrastructure: becoming locked into an unsustainable future. **Sustainability**, v. 14, n. 8, p. 4829, 2022.

17 REGILME, Salvador Santino F. Artificial intelligence colonialism: Environmental damage, labor exploitation, and human rights crises in the Global South. **SAIS Review of International Affairs**, v. 44, n. 2, p. 75-92, 2024.

18 FAYZA, Farbin et al. Photonics for sustainable AI. **Communications Physics**, v. 8, n. 1, p. 403, 2025.

19 VURUMA, Sai Krishna Revanth et al. From cloud to edge: rethinking generative AI for low-resource design challenges. **arXiv preprint arXiv:2402.12702**, 2024.

um large model a um small model. Neste sentido, os autores apontam que os smalls model são alternativas por serem capazes de reduzir demandas energéticas e a dependência de data centers de hiperescala e conectividade contínua (menor uso de internet, porque o processamento pode ocorrer localmente em dispositivos "na ponta", *edge*).

Outro estudo proeminente que aborda caminhos para reduzir os impactos da infraestrutura da Inteligência Artificial com base nos modelos é o de Clemm *et al.*[20], indicando o uso de *Low-Rank Adaptation* (LoRA), uma técnica aplicada a modelos de linguagem pequenos capaz de reduzir em aproximadamente 10.000 vezes o número de parâmetros envolvidos no treinamento, consequentemente, diminuindo o orçamento computacional e a quantidade de GPUs necessária. Outra alternativa é a compreensão (*pruning*) de parâmetros considerados redundantes, um exemplo é o da AlexNet: uma rede neural utilizada em tarefas de reconhecimento de imagens, que foi testada aplicação desta técnica e conseguiu comprimir o modelo em 71 vezes.

Esses exemplos servem para ilustrar como técnicas de redução e compressão podem constituir alternativas para diminuir a demanda computacional e de hardware aos sistemas de IA e, potencialmente, seus impactos ambientais.

Entretanto é importante dimensionar que a adoção das técnicas citadas, apesar de ser uma alternativa viável para a adoção de modelos cujo resultado é útil a uma atividade social específica, não implica numa correlação direta com resultados eticamente desejados[21]. As técnicas também não respondem ao dilema proposto no Paradoxo de Jevons, mencionado por autoras como Luccioni, Strubell e Crawford[22], no qual o aumento da eficácia média de uma tecnologia, no caso os modelos de IA, implicaria numa adoção global que traria de volta o problema do seu custo ambiental[23]. Por ser mais eficiente, seria mais adotada, por ser mais adotada, o cômputo global seria de predação ecológica e inadequação às diretrizes de proteção ambiental.

---

20 CLEMM, Christian et al. Towards Green AI: Current status and future research. In: 2024 **Electronics Goes Green 2024+(EGG)**. IEEE, 2024. p. 1–11.

21 PAULA, E.; SONI, J.; UPADHYAY, H. et al. Comparative analysis of model compression techniques for achieving carbon efficient AI. **Scientific Reports**, v. 15, art. 23461, 2025. DOI: https://doi.org/10.1038/s41598-025-07821-w. Disponível em: https://www.nature.com/articles/s41598-025-07821-w. Acesso em: 13 out. 2026

22 LUCCIONI, Alexandra Sasha; STRUBELL, Emma; CRAWFORD, Kate. From efficiency gains to rebound effects: the problem of Jevons' paradox in AI's polarized environmental debate. In: **ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FAccT '25)**, 2025, Atenas. Proceedings [...]. New York: Association for Computing Machinery, 2025. p. 76–88. DOI: https://doi.org/10.1145/3715275.3732007. Disponível em: https://facctconference.org/static/docs/facct2025-206archivalpdfs/facct2025-final37-acmpaginated.pdf. Acesso em: 13 out. 2026.

23 LUNA CARMENO, Noemi; DOMINGOS, Tiago; O'NEILL, Daniel W. The impacts of artificial intelligence on environmental sustainability and human well-being. arXiv preprint arXiv:2602.24091, 2026.

4. Deficiência de evidências científicas com recortes geográficos do nordeste brasileiro



Segundo o levantamento bibliográfico feito no escopo 3 do presente trabalho, notamos um número significativamente menor de publicações científicas disponíveis sobre os impactos socioambientais da infraestrutura da IA com o recorte geográfico específico do Nordeste brasileiro. Além da literatura encontrada ser majoritariamente em inglês, com uma quantidade mais reduzida para filtragens em português, as pesquisas específicas centralizadas no Nordeste do país são nulas para os anos entre 2022 e 2025. Ao ampliar a delimitação de pesquisa para o cenário nacional (escopo 4), identificamos um aumento no número de resultados. Isso evidencia que a discussão sobre os impactos ambientais associados aos modelos de IA e a pesquisa científica não abrangem a região.

No entanto, ainda que haja uma invisibilização do debate acerca dos impactos socioambientais da infraestrutura da IA no Nordeste, isso não significa que os impactos sejam insignificantes, e sim que a atual conjuntura da região no contexto tecnológico não é suficientemente estudada. Nem mesmo após uma dupla checagem nos repositórios institucionais das universidades UFPE, UFRPE e UPE foram encontrados trabalhos a respeito disso.

A infraestrutura da IA vem se expandindo pelo Nordeste, uma região historicamente marcada por clima semiárido e secas prolongadas, sem que exista uma expansão proporcional da pesquisa científica acerca de seus impactos socioambientais para a região. O Ceará ocupa a primeira posição entre os estados com maior número de data centers do Nordeste, com 11 infraestruturas instaladas do total de 19 situadas no Nordeste. Conforme afirmado pelo atual secretário de Ciência, Inovação e Desenvolvimento Tecnológico de Caucaia (Ceará), há ao menos 4 data centers confirmados na região, além da construção do data center da *ByteDance*, todos com potencial energético acima de 150 MW, ou seja, data centers de hiperescala.

Destaca-se que o município de Caucaia, no Ceará, tornou-se estratégico para implementação de infraestrutura de IA, devido a presença de cabos submarinos de fibra óptica e por seu potencial de energias renováveis: energia eólica e solar. Segundo Brochado Neto e Kazak, o projeto para implementação de um data center de grande escala na cidade foi projetado para ser o maior da América Latina, com

uma área de 700.000 metros quadrados.[24] No entanto, a crise hídrica enfrentada pelos moradores da região tem sido ignorada pelas autoridades locais, que enxergam nela um potencial de protagonismo tecnológico.

Importante destacar que o Nordeste já é historicamente marcado por cenários de injustiça socioambiental e eventos climáticos extremos[25]. A região é a que mais sofre com desastres climáticos no Brasil, como é o caso das secas, e o próprio estado do Ceará está entre aqueles com maior volume populacional em situação de risco socioambiental.

No entanto, os projetos de implementação de infraestruturas de IA contribuem para intensificar injustiças socioambientais e climáticas severas, reforçando violências históricas contra indivíduos ou grupos sociais mais vulneráveis[26], as que são afetadas pelos impactos ambientais atrelados à implementação de infraestruturas de IA, especialmente aquelas de larga escala.[27] Em Caucaia, as populações indígenas já sentem os efeitos da inovação tecnológica acelerada, e temem que o consumo de energia e água para manutenção dessas estruturas agrave a crise hídrica local, impactando diretamente o cotidiano dos residentes. [28]

Em Sergipe, o projeto de implementação de data center de hiperescala prevê um consumo mais elevado de energia elétrica do que o total consumido pelo estado inteiro. No documentário ***Vampiros de recursos naturais***, os moradores do município de Nossa Senhora do Socorro relatam que a região encontra-se constantemente em crise hídrica e elétrica, fator que se intensificará com a instalação de uma infraestrutura de IA de hiperescala, que se sustenta pelo consumo de água e energia. [29]

---

24 BROCHADO NETO, D.A., KAZAK, R. Environmental risks of data center infrastructure in Northeastern Brazil from an ecocide risk perspective. **Discov Sustain** 7, 1472 (2026). DOI: https://doi.org/10.1007/s43621-026-04462-5.

25 KNOX, Winifred; FERREIRA, José Gomes. Desastre ambiental e zonas de sacrifício: o derramamento de petróleo no Nordeste do Brasil e políticas públicas de Estado. Revista da **ANPEGE**, v. 18 n. 37, p. 89-105, (2022). https://doi.org/10.5418/ra2022.v18i37.16273

26 FURTADO, Fabrina. De zonas de sacrifício à luta por justiça ambiental: a cidade é feita pra quem? **Coletiva**, Recife, n. 24, fev./mai. (2019). Disponível em: https://www.coletiva.org/dossie-direito-a-cidade-n24-artigo-de-zonas-de-sacrificio-a-lutas-por-justica-ambiental.

27 TEIXEIRA, Rylanneive Leonardo Pontes et al. Desastres socioambientais e injustiças climáticas no Nordeste do Brasil. **Revista de Desenvolvimento Regional**, v. 22 n. 1, jan./mar. (2025). https://doi.org/10.26767/coloquio.222025.3603.

28 MARTINS, Laís. CHIAVERINI, Tomás. 'ÁGUA PARA O POVO, NÃO PARA DATA CENTERS' Moradores se unem para barrar data center do TikTok no Ceará, **Intercept Brasil**, 2025. https://www.intercept.com.br/2025/09/17/moradores-se-mobilizam-contra-data-center-tiktok-ceara/

29 WENDAL, Carmo. "Vampiros de recursos naturais": data centers previstos para Sergipe devem consumir mais energia do que todo o estado. **Mangue Jornalismo**, 2026. https://manguejornalismo.org/vampiros-de-recursos-naturais-data-centers-previstos-para-sergipe-devem-consumir-mais-energia-do-que-todo-o-estado/

Brochado Neto e Kazak (2026) estabelecem categorias de risco ambiental relacionadas ao desenvolvimento e operação de data centers em Caucaia, com destaque para: **1) risco energético**, visto que os moradores já sofrem com cortes de energia frequentes; **2) risco hídrico**, em uma área que já sofre com escassez de água regularmente; **3) pegada de carbono**, devido à necessidade de complementação da rede elétrica no Brasil, dependente de usinas termelétricas. Considera-se, ainda, os riscos para a saúde da população local, exposta à poluição do solo e da água pela liberação de substâncias químicas provenientes da infraestrutura da IA, além da poluição sonora. Dessa forma, mesmo que se venda a ideia de uma IA sustentável, que utiliza energias renováveis, ainda é uma cadeia dependente da emissão de Gases de Efeito Estufa (GEE).

Ainda que o Nordeste, especialmente o Ceará, tenha se tornado um polo de data centers no Brasil, a literatura não possui quase nenhuma pesquisa científica que evidencie os impactos socioambientais atrelados à infraestrutura da IA nesta localidade [30]. Isso evidencia um aspecto importante da pesquisa científica sobre a infraestrutura da IA: um apagamento dos impactos socioambientais no Nordeste.

No intervalo escolhido para a presente pesquisa (2022-2025), não foram encontradas obras para os argumentos de busca utilizados especificamente para o Nordeste. No entanto, o texto de Brochado Neto (2025)[31] foi encontrado através de buscas posteriores, utilizando os termos Nordeste + impactos ambientais + data centers.

Assim, sustenta-se uma estrutura de opacidade ambiental, que integra a infraestrutura da IA[32] e visa ocultar os impactos socioambientais atrelados ao seu funcionamento, como esgotamento de recursos hídricos, emissão de carbono e consumo de energia elevado, além dos impactos diretos à saúde da comunidade local devido ao contato direto de substâncias tóxicas com o solo e a água da região afetada[33]. Dessa forma, cria-se, socialmente, um ideal tecnológico de que a IA não é uma estrutura física, mas que está na nuvem, tornando opacos aos olhos os impactos presentes em todo seu ciclo de vida, da extração de metais ao descarte de lixo eletrônico.

---

30 LUNA CARMENO, Noemi; DOMINGOS, Tiago; O'NEILL, Daniel W. The impacts of artificial intelligence on environmental sustainability and human well-being. arXiv preprint arXiv:2602.24091, 2026.

31 BROCHADO NETO, Djalma Alvarez. DATA CENTERS NO CEARÁ: AS FRONTEIRAS ENTRE A REVOLUÇÃO TECNOLÓGICA E O 'RISCO ECOCÍDIO'. **Revista Juridicidade Constitucional e Democracia**, v. 3 n. 5 (2025). https://doi.org/10.59776/2965-3290.2025.7630

32 FERNANDES, André, et al. OPACIDADE AMBIENTAL NA DISCUSSÃO SOBRE TRANSPARÊNCIA DA ECONOMIA DA INTELIGÊNCIA ARTIFICIAL. **Instituto de Pesquisa em Direito e Tecnologia do Recife - IP.rec**, junho de 2025.

33 ZHUK, op. cit.

Ademais, é importante destacar a dinâmica de colonialismo digital[34] presente na estrutura econômica da IA e em sua cadeia produtiva, em que o custo ambiental está concentrado em países do Sul Global para atender a uma demanda de ascensão tecnológica do exterior, utilizando nossos recursos naturais e fontes de energia renováveis para alimentar o funcionamento dessas infraestruturas.

Além da fonte de energia renovável e dos cabos submarinos de fibra óptica presentes no Nordeste, a falta de regulação ambiental para data centers também torna o Brasil um país estratégico para investimento de empresas estrangeiras na construção de infraestruturas tecnológicas da IA. Essa carência regulatória torna a região suscetível a explorações irregulares, sem que haja transparência referente aos impactos ambientais inerentes.

Portanto, a ausência de pesquisas científicas relacionadas aos impactos socioambientais da infraestrutura da IA, focada em *small models* e *large models* e centralizada na região Nordeste, demonstra que existe pouca visibilidade sobre o assunto, ainda que a exploração dessa região continue avançando rapidamente. O avanço tecnológico tem sido desproporcional ao avanço regulatório, aumentando a vulnerabilidade de populações locais à violação de direitos, bem como contribuindo para uma exploração ambientalmente irresponsável das energias renováveis do Nordeste brasileiro.

34 FAUSTINO, Deivison; LIPPOLD, Walter. **Colonialismo digital**: por uma crítica hacker-fanoniana. São Paulo: Boitempo, 2023.



Conclusão



De acordo com o debate resultado da revisão de literatura feita, podemos observar que o ponto crucial de convergência das possibilidades de mitigação dos impactos ambientais devido a infraestrutura da IA depende de uma reestruturação do seu ciclo de vida, o que compreende as fases de: 1) fabricação (preferência) dos componentes utilizados, 2) processos de treinamento, 3) operação do modelo. Neste sentido, *small models* surgem como alternativa frutífera devido a otimização de sua capacidade operacional, capacitando, por conseguinte, a redução dos impactos socioambientais.

A otimização do poder computacional é uma das soluções possíveis para gerar a redução dos impactos ambientais referentes à infraestrutura da IA. Além disso, a utilização de técnicas como a poda, reduzindo o tamanho do modelo e acelerando a capacidade computacional, o aumento de dados para diminuição do consumo de energia, bem como a quantização, que reduz a precisão numérica dos parâmetros também podem contribuir para diminuição dos custos ambientais elevados.

Por si só, a utilização de fontes de energia renováveis não torna a infraestrutura da IA sustentável. Nos casos do Nordeste expostos, Sergipe e Ceará, nota-se que a exploração de energias renováveis traz impactos socioambientais significativos ao meio ambiente e às comunidades locais e povos tradicionais. A escassez hídrica, o consumo elevado de energia e a produção de ruído[35] impactam diretamente o cotidiano desses indivíduos, que são submetidos reiteradamente a contextos de injustiça socioambiental para sustentar o funcionamento dessas infraestruturas.

Nesse sentido, é necessário estabelecer critérios de transparência e o monitoramento constante dos impactos ambientais atrelados à infraestrutura da IA, utilizando métricas de sustentabilidade ambiental e exigindo declarações de impacto referentes às emissões de carbono, ao consumo de água e energia, bem como ao descarte de resíduos provenientes dela.

---

35 WHEELER, Megan Rand et al. Infrasound and low-frequency noise in data center environments: a narrative review toward health-protective acoustic design standards. **Clean Technologies**, v. 8 n. 4, ago. (2026). https://doi.org/10.3390/cleantechnol8040126

No entanto, a carência regulatória é um problema central para o efetivo controle de danos ambientais no Nordeste. Ao mesmo tempo que a região tem se tornado cada vez mais um polo tecnológico para implementação de *data centers*, a regulação não tem se desenvolvido proporcionalmente, tornando o território e os moradores locais vulneráveis à injustiças socioambientais.

Por fim, a ausência de pesquisas sobre impactos socioambientais com base nos modelos de linguagem no Nordeste é parte da estrutura de opacidade ambiental inerente à economia da IA. Portanto, destaca-se a importância de investimento em pesquisa científica de otimização computacional de *small models*, bem como de impactos socioambientais da infraestrutura da IA, com recorte geográfico específico para região do Nordeste.

Tal achado demonstra uma deficiência importante não apenas em termos de produção científica em sentido estrito, mas na capacidade de construção de políticas públicas baseadas em evidências. Fica demonstrado, portanto, que os projetos de implementação de centros de dados atualmente em marcha na região carecem de base científica orientativa a dar sustentação ao seu desenvolvimento ecologicamente orientado.

Tal ausência indica um risco elevado de violações socioambientais e uma atuação do Poder Público desamparada de indicadores científicos. Do ponto de vista da participação pública significativa, os achados deste trabalho parecem reforçar a prática cotidiana do IP.rec que verifica uma completa ausência de balizas democráticas na construção das políticas do setor no Brasil, com ausência cabal de participação da academia e da sociedade civil. O presente relatório tenta ser uma contribuição, mesmo que inicial, para esta problemática.


OTIMIZAÇÃO DO PODER COMPUTACIONAL
Reduz a quantidade de operações, memória e recuros necessários para reinar e executar o modelo
PODA
Remove parâmetros redundantes
QUANTIZAÇÃO
32 → 8
Reduz a precisão numérica dos parâmetros
DESTILAÇÃO DE CONHECIMENTO
Transfere conhecimento de um modelo maior para um menor
ADAPTAÇÃO DE BAIXA ORDEM (LoRA)
Atualiza apenas parte dos parâmetros
MODELOS MENORES
Atendem certas tarefas com menos recursos (inclusive localmente)
Resultado: menos operações, menor uso de memória e menor demanda por hardware.
MECANISMOS FÍSICOS E MODOS DE TREINAMENTO
Aumentam a eficiência a infraestrutura e do processo de desenvolvimento da IA
HARDWARE MAIS EFICIENTE
Aceleradores e equipamentos com maior eficiência energética
SOFTWARE E ALGORITMOS OTIMIZADOS
Arquiteturas e métodos de treinamento mais eficientes
AMOSTRAGEM E AUMENTO DE DADOS
Reduzem o custo computacional para obter desempenho equivalente
OTIMIZAÇÃO DE INFERÊNCIA
Respostas mais eficientes às solicitações dos usuários
DATA CENTERS MAIS EFICIENTES
Melhor gestão de energia e refrigeração
USO DE FONTES RENOVÁVEIS
Eleticidade com menores emissões de gases de efeito estufa
AMPLIAÇÃO DA VIDA ÚTIL DO HARDWARE
Reduz a substituição de equipamentos e o lixo eletrônico
Resultado: menos consumo de energia, água materiais ao longo do ciclo de vida.
POTENCIAL DE REDUÇÃO DO IMPACTO AMBIENTAL
A cominação das técnicas e dos mecanismos pode diminuir os impactos ambientais da IA.
Menor consumo de eletricidade
Menores emissões de gases de efeito estufa
Menor uso de água (para resfriamento)
Menor extração de minerais (pela maior vida útil dos equipamentos e menor demanda por hardware)
Menos resíduos eletrônicos
Resultado: menos consumo de energia, água materiais ao longo do ciclo de vida.
Fatores que influenciam os resultados
Escala de uso e demanda total
Aplicação e contexto
Matriz energética e localização dos data centers
Políticas, governança e regulação
Impactos avaliados
Energia
Emissões de GEE
Água
Recursos minerais
Resíduos eletrônicos
IMPACTO AMBIENTAL











ip.rec

Instituto de Pesquisa em Direito & Tecnologia do Recife

acesse **ip.rec.br**

@ip.rec · @institutoiprec · @ip.rec · LinkedIn/company/institutoiprec