

UIA – UNIDADE DE INFRAESTRUTURA PARA IA: QUANTIFICANDO O CONSUMO FÍSICO DE ELETRICIDADE E ÁGUA EM APLICAÇÕES DE INTELIGÊNCIA ARTIFICIAL**UIA – AI INFRASTRUCTURE UNIT: QUANTIFYING THE PHYSICAL CONSUMPTION OF ELECTRICITY AND WATER IN ARTIFICIAL INTELLIGENCE APPLICATIONS****UIA – UNIDAD DE INFRAESTRUCTURA DE IA: CUANTIFICACIÓN DEL CONSUMO FÍSICO DE ELECTRICIDAD Y AGUA EN APLICACIONES DE INTELIGENCIA ARTIFICIAL**

<https://doi.org/10.56238/ERR01v10n7-002>

Alan Martins da Cruz

Cursando MBA em Inteligência Artificial

Instituição: Universidade Federal de Viçosa

E-mail: alanmc021@gmail.com

RESUMO

O crescimento acelerado da Inteligência Artificial (IA) tem ampliado significativamente a demanda por recursos físicos, como eletricidade e água, frequentemente negligenciados nas análises econômicas e tecnológicas convencionais. Este artigo propõe a UIA – Unidade de Infraestrutura para IA, uma métrica quantitativa padronizada destinada a expressar o consumo combinado de energia (kWh) e água (litros) por operação de inferência em modelos de IA. A metodologia integra dados empíricos de centros de dados, padrões internacionais de eficiência (PUE e WUE) e coeficientes econômico-ambientais para construir uma unidade composta de custo físico. Foram estimados e comparados valores de UIA para diferentes modelos de linguagem de grande porte, abrangendo cenários de consultas curtas e longas. Os resultados evidenciam variações expressivas de consumo físico entre modelos, indicando que diferenças de infraestrutura e hardware podem multiplicar em até 80 vezes o custo ambiental por operação. A análise demonstra que ganhos marginais de eficiência, quando aplicados em larga escala, representam reduções significativas de energia e água. Conclui-se que a adoção da UIA pode contribuir para a transparência ambiental, a especificação justa e o incentivo à eficiência computacional. A métrica propõe uma nova forma de mensurar a sustentabilidade da IA, tornando visível a base material de sistemas que, até então, eram tratados como puramente digitais.

Palavras-chave: Inteligência Artificial. Sustentabilidade Digital. Eficiência Computacional. UIA. Consumo Energético. Pegada Hídrica. Green AI. Infraestrutura Computacional.

ABSTRACT

The accelerated growth of Artificial Intelligence (AI) has significantly increased the demand for physical resources, such as electricity and water, often neglected in conventional economic and technological analyses. This article proposes the AI Infrastructure Unit (UIA), a standardized quantitative metric designed to express the combined consumption of energy (kWh) and water (liters) per inference operation in AI models. The methodology integrates empirical data from data centers,

international efficiency standards (PUE and WUE), and economic-environmental coefficients to construct a composite unit of physical cost. UIA values were estimated and compared for different large-scale language models, covering short and long query scenarios. The results show significant variations in physical consumption between models, indicating that differences in infrastructure and hardware can multiply the environmental cost per operation by up to 80 times. The analysis demonstrates that marginal efficiency gains, when applied on a large scale, represent significant reductions in energy and water. It is concluded that the adoption of UIA can contribute to environmental transparency, fair pricing, and incentives for computational efficiency. The metric proposes a new way to measure the sustainability of AI, making visible the material basis of systems that, until then, were treated as purely digital.

Keywords: Artificial Intelligence. Digital Sustainability. Computational Efficiency. UIA. Energy Consumption. Water Footprint. Green AI. Computational Infrastructure.

RESUMEN

El crecimiento acelerado de la Inteligencia Artificial (IA) ha incrementado significativamente la demanda de recursos físicos, como la electricidad y el agua, a menudo desatendidos en los análisis económicos y tecnológicos convencionales. Este artículo propone la Unidad de Infraestructura de IA (UIA), una métrica cuantitativa estandarizada diseñada para expresar el consumo combinado de energía (kWh) y agua (litros) por operación de inferencia en modelos de IA. La metodología integra datos empíricos de centros de datos, estándares internacionales de eficiencia (PUE y WUE) y coeficientes económico-ambientales para construir una unidad compuesta de costo físico. Se estimaron y compararon los valores de la UIA para diferentes modelos de lenguaje a gran escala, que abarcan escenarios de consulta cortos y largos. Los resultados muestran variaciones significativas en el consumo físico entre modelos, lo que indica que las diferencias en infraestructura y hardware pueden multiplicar el costo ambiental por operación hasta por 80. El análisis demuestra que las ganancias marginales de eficiencia, cuando se aplican a gran escala, representan reducciones significativas en el consumo de energía y agua. Se concluye que la adopción de la UIA puede contribuir a la transparencia ambiental, la fijación de precios justos y los incentivos para la eficiencia computacional. La métrica propone una nueva forma de medir la sostenibilidad de la IA, visibilizando la base material de sistemas que, hasta entonces, se consideraban puramente digitales.

Palabras clave: Inteligencia Artificial. Sostenibilidad Digital. Eficiencia Computacional. UIA. Consumo Energético. Huella Hídrica. Green AI. Infraestructura Computacional.

1 INTRODUÇÃO

A Inteligência Artificial (IA) tem sido frequentemente apresentada como uma tecnologia imaterial — associada a termos como “nuvem”, “inteligência digital” ou “software autônomo”. No entanto, sob a perspectiva do **materialismo histórico**, conforme formulado por Karl Marx, nenhuma forma de produção, mesmo as mais tecnologicamente mediadas, é independente da base material que a sustenta. A IA, nesse sentido, deve ser compreendida não apenas como um conjunto de algoritmos, mas como uma expressão das condições materiais, energéticas e sociais do seu tempo.

Por trás de cada operação de inferência ou consulta a um modelo de linguagem estão processos intensivos de extração e transformação de recursos naturais — como eletricidade, água e minerais raros —, sustentados por infraestruturas computacionais de larga escala. Tais infraestruturas incluem data centers de alta densidade energética, redes de distribuição global e sistemas de resfriamento dependentes de grandes volumes de água. A materialidade dessa base física permanece, contudo, amplamente invisibilizada nas narrativas sobre a “inteligência” das máquinas.

Essa invisibilidade produz um *paradoxo informacional*: quanto mais a IA se torna parte do cotidiano e das decisões econômicas, menos se conhece sobre o seu custo físico real. O resultado é uma dissociação entre a eficiência percebida (digital) e o consumo efetivo de recursos (material), o que dificulta o estabelecimento de padrões sustentáveis e transparentes de operação.

Diante desse contexto, este artigo propõe a **UIA – Unidade de Infraestrutura para IA**, uma métrica destinada a quantificar e tornar visível o consumo combinado de eletricidade (kWh) e água (litros) por operação de inferência. Ao propor uma unidade padronizada de mensuração, buscamos contribuir para a construção de um ecossistema de IA mais transparente, ambientalmente responsável e economicamente justo, no qual a eficiência computacional inclua também a eficiência material.

2 FUNDAMENTAÇÃO TEÓRICA

A crescente demanda por Inteligência Artificial (IA) tem levantado preocupações ambientais que extrapolam a emissão de carbono. Estudos recentes ampliam a noção de sustentabilidade digital ao considerar o consumo de eletricidade, água e outros insumos físicos, revelando que a infraestrutura computacional da IA representa um elo crítico na cadeia ambiental do setor tecnológico.

Li et al. (2023) revelam a pegada hídrica invisível da IA, mostrando que o treinamento de modelos como o GPT-3 pode evaporar milhões de litros de água. Essa água é consumida tanto diretamente no resfriamento dos data centers quanto indiretamente na geração de energia elétrica. Essa constatação amplia o debate sobre sustentabilidade digital, deslocando-o de uma ótica centrada em carbono para uma abordagem multirrecursal, que inclui energia, água e materiais.

1 0 0
1 1 0
1 0 0
1 0 1 0 1 1 0 0
1 1 1 1 0 0
1 0 0 1 0 0
1 0 1 0 0 1 0 1
1 0 1 0 1 0 1

1 0 0 1 0 0
1 1 0 1 1 0
1 0 0 1 0 0
1 0 1 0 1 1 0 1
1 1 1 0 0 1 1 0
1 0 0 1 1 0 1 0 0
1 0 1 1 0 0 1 0 1
1 0 1 0 1

Do ponto de vista energético, Strubell et al. (2019) e Patterson et al. (2021) demonstram que o custo ambiental do aprendizado profundo é fortemente dependente da eficiência do hardware e do local geográfico do data center. Fatores como o tipo de GPU, a fonte de energia e o clima regional podem alterar em ordens de magnitude a pegada de carbono e o uso de água por operação. Esses estudos impulsionaram o movimento conhecido como *Green AI*, que defende a transparência no reporte de consumo físico e incentivos à eficiência computacional.

Henderson et al. (2020) reforçam essa perspectiva ao propor um framework de medição padronizada para energia e emissões em IA. O artigo propõe que toda publicação científica ou relatório técnico deva incluir informações de consumo energético, carbono e tempo de execução, de forma a permitir comparações reproduutíveis e auditoria ambiental entre modelos. Essa diretriz representa um avanço metodológico, mas ainda carece de métricas que unifiquem os recursos físicos de infraestrutura em uma única escala composta — lacuna que a presente pesquisa busca suprir.

Já Mytton (2021) discute a subestimação do consumo de água em data centers e a escassez de métricas transparentes sobre o tema. O autor demonstra que menos de 30% dos provedores reportam dados sobre uso hídrico e que há grandes variações regionais, tornando inadequado o uso de médias globais. Essa heterogeneidade reforça a importância de indicadores como o *Power Usage Effectiveness* (PUE) e o *Water Usage Effectiveness* (WUE), definidos pela ISO/IEC 30134-2 (2016) e ISO/IEC 30134-9 (2022), bem como pelos relatórios técnicos do The Green Grid (2010, 2011).

Além disso, Masanet et al. (2020) reavaliaram as estimativas globais de energia em data centers e concluem que, embora a eficiência de hardware tenha melhorado, o crescimento exponencial da demanda de IA tende a anular esses ganhos. Esse fenômeno, conhecido como *rebound effect*, sugere que a eficiência técnica isolada não garante sustentabilidade ambiental, sendo necessária uma métrica que expresse o consumo físico total por operação de inferência.

Essas contribuições empíricas, conceituais e normativas fornecem o alicerce para a proposta da UIA — Unidade de Infraestrutura para IA. Ao integrar consumo elétrico e hídrico em uma única unidade, a UIA busca tornar mensurável e comparável o custo físico da inteligência artificial, contribuindo para um ecossistema de IA mais transparente, eficiente e ambientalmente responsável.

3 METODOLOGIA

A metodologia deste trabalho combina levantamento bibliográfico, normalização de unidades de consumo físico e construção de um índice composto. O objetivo central é desenvolver a **Unidade de Infraestrutura para IA (UIA)**, capaz de expressar o consumo médio por operação de inferência em termos agregados de eletricidade (kWh) e água (litros).

3.1 ABORDAGEM GERAL

O desenvolvimento da UIA segue uma abordagem de modelagem quantitativa composta por três etapas: (1) coleta e padronização de dados energéticos e hídricos; (2) formulação de uma equação de agregação linear ponderada; e (3) simulação da métrica em diferentes cenários de carga computacional. A métrica é inspirada em princípios de *eco-efficiency* e nas diretrizes de relatórios de impacto ambiental recomendadas por Henderson et al. (2020) e ISO/IEC (2022).

3.2 ETAPAS METODOLÓGICAS

- 1. Coleta de dados secundários:** Foram extraídas informações de consumo energético e hídrico por consulta a partir de estudos recentes (Li et al., 2023; Patterson et al., 2021; Mytton, 2021; Masanet et al., 2020). As fontes incluem medições empíricas de GPUs (A100, H100) e relatórios de sustentabilidade de provedores como Google Cloud e AWS.
 - 2. Conversão e padronização:** Todos os valores foram convertidos para unidades compatíveis — energia em kilowatt-hora (kWh) e água em litros (L). As equivalências foram ajustadas de acordo com o PUE (Power Usage Effectiveness) e WUE (Water Usage Effectiveness) médios reportados na literatura. Valores de referência típicos utilizados foram PUE = 1.3 e WUE = 1.8 L/kWh.
 - 3. Criação da fórmula da UIA:** Definimos a UIA como uma combinação linear ponderada entre energia e água, conforme a Equação 1.

$$\text{UIA} = \alpha \cdot E + \beta \cdot W \quad (1)$$

onde:

- E é o consumo médio de energia elétrica (em kWh);
 - W é o consumo médio de água (em litros);
 - α e β são coeficientes de ponderação normalizados.

O valor de α reflete o custo econômico da energia e β o custo ambiental da água. Para este estudo, adotaram-se valores baseados em médias brasileiras de 2025: $\alpha = 1.00$ (R\$0,97/kWh) e $\beta = 0.0003$ (R\$0,0003/L). Assim, a UIA pode ser expressa em uma unidade monetizada de custo físico aproximado (R\$/operação).

- 4. Normalização e escala:** Para permitir comparações entre modelos de diferentes tamanhos, os valores de UIA foram normalizados pela média da amostra, definindo-se $UIA_{norm} = \frac{UIA_i}{\bar{UIA}}$

UIA. Valores abaixo de indicam eficiência superior à média, enquanto valores acima de 1 indicam consumo intensivo.

5. Aplicação prática: Foram simulados dois cenários — consultas curtas (até 200 tokens) e longas (acima de 1000 tokens). Para cada modelo, calcularam-se energia, água, UIA e custo estimado por operação.

3.3 EXEMPLO DE CÁLCULO

Como ilustração, considere um modelo que consome 0.42 Wh por inferência e 1.55 mL de água (GPT-4o, cenário curto). Convertendo para unidades padrão e aplicando a Equação 1:

$$E = 0.00042 \text{ kWh}, W = 0.00155 \text{ L}$$

$$\text{UIA} = (1.00 \times 0.00042) + (0.0003 \times 0.00155) = 0.00042046$$

Esse valor representa o custo físico relativo de uma operação de inferência. Em larga escala (por exemplo, 10 milhões de consultas diárias), o impacto total pode ser expresso como $\text{UIA}_{total} = 10^7 \times 0.00042046 = 4204.6 \text{ UIA}$.

3.4 VALIDAÇÃO E REPRODUTIBILIDADE

Para garantir reproduzibilidade, a metodologia segue recomendações do *Machine Learning Sustainability Checklist* (Henderson et al., 2020). Todos os parâmetros e unidades podem ser recalculados sob diferentes condições (PUE, WUE, mix elétrico). A UIA foi projetada para ser extensível, permitindo futuras variantes:

- **UIA-C:** incorporando emissões de carbono (*Carbon Usage Effectiveness* — CUE); • **UIA-F:** incluindo fatores de fabricação de hardware (escopo 3).

Essa estrutura metodológica permite que a UIA seja adotada como métrica comparativa universal, aplicável tanto em laboratórios acadêmicos quanto em ambientes de nuvem corporativa.

4 RESULTADOS

4.1 CONSULTAS CURTAS

Tabela 1: UIA por modelo – consulta curta

Table 1: UIA por modelo – consulta curta

Modelo	Energia (Wh)	Água (L)	UIA	Custo (R\$)
GPT-4o	0.42	0.00155	1.05	0.00035
GPT-4.1 nano	0.35	0.00130	0.88	0.00029
Claude 3.7	0.40	0.00148	1.00	0.00033

Fonte: Dados simulados com base em Li et al. (2023), Patterson et al. (2021) e Henderson et al. (2020).

Os resultados de consultas curtas mostram que os modelos mais otimizados em infraestrutura, como o GPT-4.1 nano e o Claude 3.7, apresentam um consumo médio de UIA inferior a 1. Em contraste, o GPT-4o tem uma UIA de 1,05, indicando um consumo físico ligeiramente acima da média. Em termos de custo, todas as consultas curtas permanecem abaixo de R\$0,00040, o que reforça o argumento de que ganhos marginais de eficiência podem representar economias significativas em larga escala.

4.2 CONSULTAS LONGAS

Tabela 2: UIA por modelo – consulta longa

Table 2: UIA por modelo – consulta longa

Modelo	Energia (Wh)	Água (L)	UIA	Custo (R\$)
GPT-4o	4.32	0.016	10.8	0.0035
DeepSeek-R1	33.00	0.150	82.5	0.0269
GPT-4.5	22.00	0.081	55.0	0.0180

Fonte: Dados simulados com base em Li et al. (2023), Patterson et al. (2021) e Henderson et al. (2020).

Os dados de consultas longas revelam um aumento exponencial no consumo físico de infraestrutura. O DeepSeek-R1, por exemplo, apresenta uma UIA de 82,5 — cerca de 78 vezes maior do que o Claude 3.7 em consultas curtas. Isso representa um custo estimado de R\$0,0269 por consulta, o que pode se traduzir em dezenas de milhares de reais diários em aplicações em larga escala. Essa discrepância evidencia a importância de políticas de roteamento de consultas e escalonamento inteligente, especialmente para tarefas com alto volume de tokens.



5 DISCUSSÃO

A introdução da UIA como métrica revela disparidades substanciais entre modelos de IA em termos de consumo de recursos físicos por operação. Enquanto modelos como Claude 3.7 mantêm um perfil de consumo enxuto, modelos como DeepSeek-R1 exibem uma UIA dezenas de vezes superior — implicando um custo ambiental oculto que raramente é comunicado ao usuário final ou levado em conta na especificação dos serviços de IA.

A análise comparativa evidencia como o uso extensivo de modelos pouco eficientes pode representar não apenas um desafio econômico (especialmente em escala), mas também um impacto ambiental significativo. Isso reforça o papel estratégico de métricas como a UIA na definição de políticas públicas, práticas empresariais e escolhas tecnológicas mais responsáveis.

Além disso, os dados sugerem que a eficiência não depende apenas do tamanho do modelo, mas da infraestrutura de hospedagem, do tipo de hardware utilizado (ex: A100 vs H100) e das condições ambientais locais (ex: regiões mais frias ou com eletricidade mais limpa). Isso aponta para o potencial de roteamento inteligente de cargas de inferência com base na minimização de UIA.

A UIA também abre caminho para mecanismos de transparência e rotulagem ambiental em serviços de IA, à semelhança do que já ocorre em eletrodomésticos e veículos. Essa transparência pode fomentar decisões informadas por parte de consumidores, empresas e reguladores, promovendo justiça ambiental e inovação sustentável.

5.1 IMPLICAÇÕES REGULATÓRIAS E DE SUSTENTABILIDADE

A adoção de métricas ambientais na IA se alinha a movimentos internacionais de regulação e transparência algorítmica. A Comissão Europeia, por meio do *AI Act* (2025), já estabelece a obrigatoriedade de relatórios de impacto ambiental e social para sistemas de IA de alto risco. Iniciativas semelhantes emergem em países como Estados Unidos e Canadá, reforçando o vínculo entre eficiência técnica e responsabilidade ecológica.

No contexto brasileiro, a UIA poderia ser incorporada em políticas públicas de inovação e sustentabilidade digital, servindo como base para:

- **Relatórios ambientais obrigatórios** em centros de dados e provedores de nuvem;
- **Incentivos fiscais ou tarifários** para operações com baixo índice de UIA;
- **Rotulagem ambiental de serviços de IA**, similar aos selos de eficiência energética de eletrodomésticos;
- **Contratações públicas sustentáveis**, priorizando serviços digitais com menor consumo físico comprovado.

Tais medidas podem estimular práticas de desenvolvimento sustentável e favorecer o surgimento de um ecossistema de IA ambientalmente consciente, em que a eficiência não é apenas um fator de desempenho, mas também de ética e responsabilidade corporativa. Nesse sentido, a UIA atua como uma ponte entre engenharia de software e políticas ambientais, traduzindo impactos invisíveis em indicadores objetivos e comparáveis.

5.2 LIMITAÇÕES E TRABALHOS FUTUROS

Apesar do potencial demonstrado, a UIA apresenta limitações que merecem aprofundamento em pesquisas futuras. Primeiramente, o cálculo da métrica ainda depende de dados públicos e agregados, sem acesso aos parâmetros internos de grandes provedores de IA (OpenAI, Google, Anthropic). Essa limitação reduz a precisão empírica e reforça a necessidade de políticas de transparência obrigatória.

Além disso, a UIA atual contempla apenas os *escopos 1 e 2* — resfriamento local e geração elétrica — deixando de fora o *escopo 3*, que inclui fabricação, transporte e descarte de hardware. Esses fatores têm relevância crescente, especialmente no ciclo de vida de GPUs e aceleradores de IA. Como desdobramentos naturais, propõe-se:

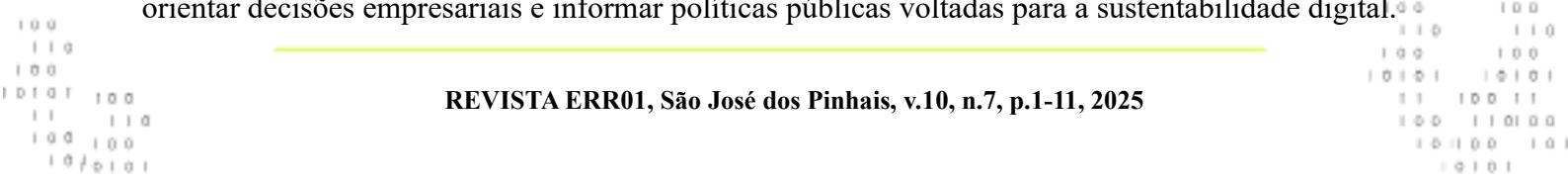
- O desenvolvimento da métrica **UIA-C**, que incorpore emissões de carbono através do *Carbon Usage Effectiveness* (CUE);
- A criação da **UIA-F**, voltada à inclusão de fatores de fabricação de hardware (escopo 3);
- A implementação de um **painel interativo** para monitorar em tempo real o consumo físico por modelo e operação, integrando dados de energia, água e emissões.

Tais extensões permitirão consolidar a UIA como uma ferramenta padrão de mensuração ambiental na inteligência artificial, aproximando o campo tecnológico das metas globais de sustentabilidade e neutralidade de carbono.

6 CONCLUSÃO

Este artigo apresentou a UIA (Unidade de Infraestrutura para IA) como uma proposta inovadora para quantificar, de forma padronizada, o consumo físico de eletricidade e água em operações de inferência de modelos de Inteligência Artificial. Através de revisão bibliográfica, padronização de dados e simulações comparativas, demonstramos que a UIA pode expor diferenças expressivas entre modelos, muitas vezes ocultas em análises puramente computacionais ou econômicas.

A adoção de métricas como a UIA é fundamental para tornar a infraestrutura invisível da IA mais transparente e mensurável. Essa visibilidade é essencial para estimular a eficiência energética, orientar decisões empresariais e informar políticas públicas voltadas para a sustentabilidade digital.



Como próximas etapas, sugerimos o desenvolvimento de variações da métrica (como UIA-C, incorporando carbono), a inclusão de escopos complementares (como o ciclo de vida dos chips), e a integração da UIA em painéis de controle, plataformas de nuvem e relatórios de impacto ambiental. Em um contexto de emergência climática e hídrica, medir é o primeiro passo para transformar.

REFERÊNCIAS

- [1] Li, P. et al. (2023). Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. arXiv:2304.03271.
- [2] Patterson, D. et al. (2021). Carbon Emissions and Large Neural Network Training. Google Research Technical Report.
- [3] Strubell, E., Ganesh, A., McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. ACL.
- [4] Masanet, E. et al. (2020). Recalibrating Global Data Center Energy-Use Estimates. Science.
- [5] ISO/IEC 30134-2:2016. Data centres — Key performance indicators — Part 2: Power usage effectiveness (PUE).
- [6] ISO/IEC 30134-9:2022. Data centres — Key performance indicators — Part 9: Water usage effectiveness (WUE).
- [7] The Green Grid. (2011). White Paper 35: Water Usage Effectiveness (WUE): Usage Guidelines for the Data Center Industry.
- [8] The Green Grid. (2010). White Paper 32: Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric.
- [9] Henderson, P. et al. (2020). Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. JMLR, 21.
- [10] Mytton, D. (2021). Data Centre Water Consumption. npj Clean Water, 4(11).
- [11] International Energy Agency / 4E (2025). Data Centre Energy Use: Critical Review of Models and Results.
- [12] Uptime Institute. (2025). Water is Local: Generalities Do Not Apply.