Research Club Scientia Vol. 4, No 3
Frecuencia: Julio - Septiembre
ISSN: 3103 - 1315
DOI: https://doi.org/10.65415/2cc4j106

# *Green AI en la educación superior: medición de la huella energética y de carbono del uso de inteligencia artificial generativa en la docencia de programación y comparación entre modelos locales y servicios en la nube*

*Green AI in higher education: measuring the energy and carbon footprint of generative artificial intelligence use in programming teaching and comparing local models with cloud services*

*Green AI nell'istruzione superiore: misurazione dell'impronta energetica e di carbonio dell'uso dell'intelligenza artificiale generativa nella didattica della programmazione e confronto tra modelli locali e servizi cloud*

**Rodrigo José Pazmiño Pérez**[1]
**E-mail:** rjpazmino@uagraria.edu.ec
**ORCID:** https://orcid.org/0009-0000-9507-6865

**Vanessa Vergara Lozano**[2]
**E-mail:** vvergara@uagraria.edu.ec
**ORCID:** https://orcid.org/0000-0002-0644-5946

**Jennifer Gabriela Salinas Sedamanos**[3]
**E-mail:** Jsalinas@defensoria.gob.ec
**ORCID:** https://orcid.org/0000-0001-6679-3494

**Liliam Stephania Garzón Reyes**[4]
**E-mail:** lgarzon@uagraria.edu.ec
**ORCID:** https://orcid.org/0009-0005-3954-4333

**Jhonny Roberto Pilco Llongo**[5]
**E-mail:** ab.jhonny.pilco@gmail.com
**ORCID:** https://orcid.org/0009-0008-6059-4046

**Correspondencia:** Rodrigo José Pazmiño Pérez[1] **Email:** rjpazmino@uagraria.edu.ec

**Artículo de investigación**

*** Recibido:** 14-08-2026 *** Aceptado:** 24-09-2026 *** Publicado:** 29-09-2026

1. Universidad Agraria del Ecuador, Guayaquil, Ecuador.
2. Universidad Agraria del Ecuador, Guayaquil, Ecuador.
3. Defensoría Pública, Guayaquil, Ecuador.
4. Universidad Agraria del Ecuador, Guayaquil, Ecuador.
5. Universidad de Guayaquil, Ecuador.

**Resumen**

La expansión de la inteligencia artificial (IA) generativa en la educación superior se ha analizado sobre todo desde sus efectos pedagógicos, mientras que su costo ambiental permanece prácticamente ausente de la investigación educativa. Este estudio tiene como objetivo medir la huella energética y de carbono del uso de IA generativa en la docencia de programación de la carrera de Computación en Línea de la Universidad Agraria del Ecuador, comparar la ejecución local de modelos abiertos con el uso de servicios en la nube y evaluar estrategias de uso que reduzcan esa huella sin afectar la calidad de las respuestas. Se adopta un diseño mixto con predominio cuantitativo en tres fases: caracterización del uso docente y estudiantil mediante cuestionario y registro de interacciones durante ocho semanas; medición directa, con la librería CodeCarbon, de la energía y las emisiones de un banco de tareas representativas ejecutadas con modelos abiertos de tres tamaños; y estimación de la huella de los servicios en la nube por escenarios, con análisis de sensibilidad. Las emisiones se calculan con el factor de emisión oficial del Sistema Nacional Interconectado del Ecuador, cuya matriz es predominantemente hidroeléctrica. La calidad de cada respuesta se valora con una rúbrica aplicada por dos jueces, lo que permite construir un índice de eficiencia entre calidad y huella. El estudio aporta un protocolo replicable de medición para instituciones de educación superior y una base empírica para formular lineamientos de uso sostenible de la IA generativa en la docencia.

**Palabras clave:** inteligencia artificial verde; huella de carbono; consumo energético; inteligencia artificial generativa; enseñanza de la programación.

**Abstract**

The expansion of generative artificial intelligence (AI) in higher education has been examined mainly through its pedagogical effects, while its environmental cost remains almost absent from educational research. This study aims to measure the energy and carbon footprint of generative AI use in programming teaching in the Online Computing program at the Agrarian University of Ecuador, to compare the local execution of open models with the use of cloud services, and to evaluate usage strategies that reduce this footprint without compromising response quality. A mixed-methods design with a dominant quantitative strand is adopted in three phases: characterization of teacher and student use through a questionnaire and an eight-week interaction log; direct measurement, with the CodeCarbon library, of the energy and emissions of a bank of representative tasks executed with open models of three sizes; and scenario-based estimation of the cloud footprint with sensitivity analysis. Emissions are calculated with the official emission factor of Ecuador's National Interconnected System, whose electricity mix is predominantly hydroelectric. The quality of each response is rated by two judges with a rubric, which allows the construction of a quality-to-footprint efficiency index. The study provides a replicable measurement protocol for higher education institutions and an empirical basis for guidelines on the sustainable use of generative AI in teaching.

**Keywords:** green artificial intelligence; carbon footprint; energy consumption; generative artificial intelligence; programming education.

**Riassunto**

L'espansione dell'intelligenza artificiale (IA) generativa nell'istruzione superiore è stata analizzata soprattutto per i suoi effetti pedagogici, mentre il suo costo ambientale resta quasi

assente dalla ricerca educativa. Lo studio si propone di misurare l'impronta energetica e di carbonio dell'uso dell'IA generativa nella didattica della programmazione del corso di Informatica online dell'Università Agraria dell'Ecuador, di confrontare l'esecuzione locale di modelli aperti con l'uso di servizi cloud e di valutare strategie d'uso che riducano tale impronta senza compromettere la qualità delle risposte. Si adotta un disegno misto a prevalenza quantitativa in tre fasi: caratterizzazione dell'uso mediante questionario e registro delle interazioni per otto settimane; misurazione diretta, con la libreria CodeCarbon, dell'energia e delle emissioni di un insieme di compiti rappresentativi eseguiti con modelli aperti di tre dimensioni; e stima per scenari dell'impronta dei servizi cloud con analisi di sensibilità. Le emissioni sono calcolate con il fattore di emissione ufficiale del Sistema Nazionale Interconnesso dell'Ecuador, la cui matrice è prevalentemente idroelettrica. La qualità di ogni risposta è valutata da due giudici con una rubrica, il che consente di costruire un indice di efficienza tra qualità e impronta. Lo studio offre un protocollo di misurazione replicabile e una base empirica per linee guida sull'uso sostenibile dell'IA generativa nella didattica.

**Parole chiave:** intelligenza artificiale verde; impronta di carbonio; consumo energetico; intelligenza artificiale generativa; didattica della programmazione.

## Introducción

La inteligencia artificial (IA) generativa se ha incorporado con rapidez a la vida universitaria. Los modelos de lenguaje de gran escala (LLM, por sus siglas en inglés) se usan para explicar conceptos, generar ejercicios, retroalimentar trabajos y, en las carreras de computación, escribir, depurar y documentar código. La literatura ha examinado con amplitud las oportunidades y los riesgos pedagógicos de esta adopción (Kasneci et al., 2023), así como su impacto específico en la enseñanza de la programación, donde la generación automática de código obliga a replantear qué y cómo se enseña (Becker et al., 2023; Denny et al., 2024). Sin embargo, este debate ha dejado en segundo plano una dimensión que ya no es marginal: el costo energético y ambiental que implica cada interacción con estos sistemas.

El problema tiene una escala que justifica atenderlo desde la educación superior. La Agencia Internacional de Energía proyecta que el consumo eléctrico de los centros de datos podría más que duplicarse hacia 2030, con la IA como uno de sus principales impulsores (International Energy Agency [IEA], 2025). Diversos análisis advierten que la demanda eléctrica asociada a la IA podría alcanzar magnitudes comparables a las de países completos si su uso continúa expandiéndose al ritmo actual (de Vries, 2023). A ello se suman otros costos menos visibles, como el consumo de agua para la refrigeración de los centros de datos (Li et al., 2023) y la huella incorporada en la fabricación del hardware (Wu et al., 2022). Bender et al. (2021) señalaron además que quienes soportan estos costos ambientales no siempre coinciden con quienes obtienen los beneficios de la tecnología, una tensión especialmente relevante para los países del Sur Global.

Frente a este escenario, las universidades ocupan una posición particular. Por un lado, son usuarias intensivas de IA generativa y promueven su adopción en la docencia; por otro, tienen la responsabilidad de formar profesionales conscientes del impacto de las tecnologías que utilizan. La Recomendación sobre la ética de la IA de la UNESCO incluye el medio ambiente y los ecosistemas entre sus ámbitos de acción política (UNESCO, 2022), y la guía de la misma organización para el uso de la IA generativa en educación e investigación insta a las instituciones a regular su adopción con criterios propios (Miao & Holmes, 2023). La Agenda 2030 aporta el

marco de referencia, en particular los Objetivos de Desarrollo Sostenible 4, 12 y 13, referidos a la educación de calidad, la producción y el consumo responsables y la acción por el clima (Naciones Unidas, 2015).

### Green AI y el giro hacia la huella de la inferencia

El estudio del costo ambiental de la IA se consolidó a partir del trabajo de Strubell et al. (2019), que cuantificó las emisiones asociadas al entrenamiento de modelos de procesamiento del lenguaje natural y recomendó reportarlas de forma sistemática. Poco después, Schwartz et al. (2020) acuñaron la distinción entre una *Red AI*, orientada a mejorar la precisión sin considerar el costo computacional, y una *Green AI*, que incorpora la eficiencia como criterio de evaluación al mismo nivel que el rendimiento. Patterson et al. (2021) demostraron que la huella de carbono depende no solo del tamaño del modelo, sino también del hardware, de la eficiencia del centro de datos y, de manera decisiva, de la intensidad de carbono de la red eléctrica donde se ejecuta. Este último factor resulta central para el presente estudio.

La primera ola de investigación se concentró en el entrenamiento, pero la masificación de los asistentes conversacionales desplazó la atención hacia la inferencia, es decir, hacia el costo de cada consulta. Luccioni et al. (2023) estimaron la huella completa de un modelo de 176 mil millones de parámetros e incluyeron su fase de despliegue, y Samsi et al. (2023) mostraron que la energía de inferencia crece con el tamaño del modelo y con la longitud de las respuestas generadas. Luccioni et al. (2024) encontraron que los modelos generativos de propósito general consumen órdenes de magnitud más energía que los modelos especializados para la misma tarea, lo que abre la pregunta de si siempre es necesario recurrir al modelo más grande disponible. En la misma línea, el informe de la UNESCO (2025) reportó que usar modelos más pequeños y especializados, acortar las instrucciones y las respuestas y aplicar compresión puede reducir de forma sustancial el consumo sin pérdidas relevantes de desempeño.

Las cifras sobre el costo de una consulta, no obstante, varían ampliamente según la metodología. Elsworth et al. (2025), con telemetría interna de la infraestructura de Google, reportaron que la consulta de texto mediana de Gemini consumía 0,24 Wh y generaba 0,03 g de $CO_2e$ en mayo de 2025, valores inferiores a muchas estimaciones públicas. Jegham et al. (2025), mediante un marco de estimación basado en la infraestructura, encontraron amplias diferencias entre modelos comerciales, especialmente en aquellos orientados al razonamiento. Esta dispersión confirma que la huella de los servicios en la nube no puede medirse directamente desde fuera y que cualquier estimación debe acompañarse de análisis de sensibilidad. Además, Luccioni et al. (2025) advierten que las ganancias de eficiencia por consulta pueden anularse por un efecto rebote, en el que el abaratamiento del uso multiplica la cantidad de consultas, un fenómeno plausible en contextos educativos donde la IA se usa a diario.

En el plano metodológico, la comunidad ha desarrollado herramientas para medir el consumo de forma reproducible. Henderson et al. (2020) propusieron un marco para el reporte sistemático de la energía y las emisiones de experimentos de aprendizaje automático; Lannelongue et al. (2021) desarrollaron una calculadora para estimar la huella de cualquier cómputo, y la librería de código abierto CodeCarbon registra el consumo de CPU, GPU y memoria y lo convierte en emisiones según la ubicación (Courty et al., 2024). La revisión sistemática de Verdecchia et al. (2023) muestra que el campo de la Green AI ha crecido con rapidez, pero con un predominio de estudios de laboratorio centrados en el entrenamiento y con escasa atención a contextos de uso reales.

### La IA generativa en la enseñanza de la programación

La enseñanza de la programación es uno de los ámbitos donde la IA generativa ha tenido mayor penetración. Los asistentes de código resuelven con solvencia los ejercicios introductorios, explican errores y proponen alternativas de solución, lo que ha llevado a replantear los objetivos de aprendizaje y las formas de evaluación (Becker et al., 2023; Denny et al., 2024). Estas tareas suelen implicar interacciones repetidas y respuestas extensas, con bloques de código, explicaciones paso a paso y sucesivas reformulaciones, justamente los factores que la literatura asocia a un mayor consumo de inferencia (Samsi et al., 2023). Por ello, la docencia de programación ofrece un escenario apropiado para estudiar la huella de la IA generativa en condiciones de uso reales y no solo en pruebas de laboratorio.

Pese a ello, la investigación en educación en computación ha abordado la IA generativa casi exclusivamente desde sus efectos en el aprendizaje, la integridad académica y la evaluación. No se identifican estudios que cuantifiquen cuánta energía y cuántas emisiones supone su uso en una asignatura concreta, ni que examinen si determinadas prácticas docentes pueden reducir ese costo sin sacrificar la utilidad pedagógica de las respuestas.

### El contexto energético ecuatoriano

La intensidad de carbono de la electricidad determina en buena medida la huella de cualquier cómputo (Patterson et al., 2021), y en este aspecto Ecuador presenta una condición singular. Su matriz eléctrica es predominantemente hidroeléctrica, lo que se refleja en un factor de emisión bajo en comparación con redes basadas en combustibles fósiles. Según el informe oficial de la Comisión Técnica de determinación de Factores de Emisión, el factor para el cálculo de huella de carbono del Sistema Nacional Interconectado fue de 0,1616 t $CO_2$-eq/MWh para la estadística de 2024 (Comisión Técnica de determinación de Factores de Emisión de Gases de Efecto Invernadero [CTFE], 2025), superior a los valores de 0,0920 y 0,1200 t $CO_2$-eq/MWh publicados para 2022 y 2023. Esta variación, asociada a la mayor participación de la generación térmica en años de menor disponibilidad hídrica, obliga a trabajar con rangos y no con un único valor.

Este contexto plantea una pregunta que la literatura internacional no ha abordado: si la inferencia ejecutada localmente en una universidad ecuatoriana, alimentada por una red de baja intensidad de carbono, puede tener una huella menor que las consultas a servicios en la nube cuyos centros de datos operan en otras redes eléctricas, y bajo qué condiciones de tamaño de modelo y tipo de tarea se cumple esa ventaja. La respuesta tiene implicaciones prácticas para las políticas institucionales de adopción de IA en universidades públicas con recursos limitados.

### Brecha, objetivos e hipótesis

La revisión permite identificar una brecha en tres niveles. En el nivel temático, los estudios de Green AI se han centrado en el entrenamiento y en pruebas de laboratorio, mientras que la investigación educativa sobre IA generativa no ha medido su costo ambiental. En el nivel contextual, no se identifican mediciones en instituciones de educación superior latinoamericanas ni análisis que incorporen el factor de emisión de una red predominantemente hidroeléctrica. En el nivel práctico, faltan evidencias sobre la relación entre la huella y la calidad de las respuestas en tareas docentes, que es la información que un docente necesita para decidir qué modelo usar y cómo formular sus consultas.

En consecuencia, el objetivo general del estudio es medir la huella energética y de carbono del uso de IA generativa en la docencia de programación de la carrera de Computación en Línea de la Universidad Agraria del Ecuador y comparar la ejecución local de modelos abiertos con el uso de servicios en la nube. Los objetivos específicos son: (a) caracterizar el patrón de uso de IA generativa de docentes y estudiantes en las asignaturas de programación; (b) medir directamente la energía y las emisiones de un banco de tareas docentes ejecutadas con modelos abiertos de distinto tamaño; (c) estimar, mediante escenarios, la huella equivalente de los servicios en la nube y compararla con la medición local; y (d) evaluar estrategias de uso que reduzcan la huella sin disminuir la calidad de las respuestas.

Se plantean tres hipótesis. H1: la energía consumida por tarea difiere significativamente según el tamaño del modelo y el tipo de tarea docente. H2: las instrucciones concisas y las respuestas de longitud acotada reducen significativamente la energía por tarea sin una disminución significativa de la calidad valorada por los jueces. H3: en las tareas de menor complejidad, las emisiones de la inferencia local calculadas con el factor de emisión ecuatoriano son inferiores a las del escenario central estimado para los servicios en la nube.

## Metodología

### Enfoque y diseño

El estudio adopta un enfoque mixto con predominio cuantitativo (Creswell & Plano Clark, 2018), organizado en tres fases secuenciales. La primera es descriptiva y caracteriza el uso real de la IA generativa en la docencia. La segunda es un experimento computacional de medición directa con modelos abiertos, cuyo diseño factorial permite contrastar H1 y H2. La tercera combina la medición local con estimaciones por escenarios para los servicios en la nube y permite contrastar H3. El componente cualitativo, de carácter complementario, corresponde a la valoración de la calidad de las respuestas por jueces expertos y al análisis de las preguntas abiertas del cuestionario (Hernández-Sampieri & Mendoza, 2018). La articulación entre fases se resume en la Tabla 1.

**Tabla 1. Fases del estudio, técnicas y productos**

| Fase | Objetivo | Técnicas e instrumentos | Producto |
|---|---|---|---|
| **1. Caracterización del uso** | Objetivo específico (a) | Cuestionario validado por expertos; registro voluntario de interacciones durante ocho semanas | Inventario de tareas y frecuencia de uso |
| **2. Medición directa** | Objetivos específicos (b) y (d) | Banco de tareas; modelos abiertos de tres tamaños; CodeCarbon; rúbrica de calidad aplicada por dos jueces | Energía, emisiones y calidad por tarea |
| **3. Estimación y comparación** | Objetivo específico (c) | Escenarios de consumo por consulta en la nube; análisis de sensibilidad | Comparación local frente a nube |
| **Síntesis** | Objetivo general | Índice de eficiencia calidad-huella; integración de resultados | Lineamientos de uso sostenible |

*Nota. Elaboración propia a partir del diseño del estudio.*

### Contexto, población y muestra

El estudio se desarrolla en la carrera de Computación en Línea de la Universidad Agraria del Ecuador, durante un periodo académico de 2026. La población está conformada por los

docentes que imparten las asignaturas de Fundamentos de Programación y Lenguaje de Programación y por los estudiantes matriculados en ellas. Para los docentes se trabaja con la totalidad del grupo, dado su tamaño reducido. Para los estudiantes se aplica un muestreo no probabilístico por conveniencia, con participación voluntaria de los paralelos disponibles. Los criterios de inclusión son estar matriculado o asignado a una de las asignaturas, haber usado alguna herramienta de IA generativa durante el periodo y firmar el consentimiento informado. Se excluyen los registros incompletos o con menos de cuatro semanas de participación.

### Fase 1: caracterización del uso

Se aplica un cuestionario elaborado para el estudio con cuatro secciones: herramientas utilizadas, frecuencia de uso, tipos de tarea y percepción sobre el impacto ambiental de la IA. Su validez de contenido se establece mediante el juicio de cinco expertos en tecnología educativa y en ciencias de la computación, con el coeficiente V de Aiken y un criterio mínimo de 0,80 por ítem. La confiabilidad de las escalas se estima con los coeficientes alfa de Cronbach y omega de McDonald en una prueba piloto con estudiantes de otra carrera.

De forma complementaria, los participantes llevan durante ocho semanas un registro semanal de sus interacciones en un formulario breve, en el que indican la herramienta, el tipo de tarea, el número aproximado de consultas y la extensión de las respuestas. El registro no recoge el contenido de las conversaciones ni datos que permitan identificar a las personas. A partir de esta información se construye el inventario de tareas típicas que alimenta el banco de la Fase 2 y se estima el volumen semanal de consultas por participante, dato necesario para extrapolar la huella al conjunto de las asignaturas.

### Fase 2: medición directa con modelos abiertos

Con base en el inventario de la Fase 1 se construye un banco de 50 consignas distribuidas en cinco categorías de tarea docente, con diez consignas por categoría (Tabla 2). Cada consigna se formula en dos versiones: una extensa, que reproduce el estilo habitual observado en el registro, y una concisa, que conserva la misma intención con la menor cantidad de texto posible. Las consignas son revisadas por los dos jueces antes de la medición para asegurar que ambas versiones solicitan lo mismo.

**Tabla 2. Categorías del banco de tareas docentes de programación**

| Categoría | Descripción | Ejemplo de consigna |
|---|---|---|
| **Explicación conceptual** | Explicar un concepto del sílabo con un ejemplo | Explica qué es una estructura repetitiva while con un ejemplo en Python |
| **Generación de ejercicios** | Crear ejercicios con enunciado y solución | Genera tres ejercicios de funciones con ingreso de datos y su solución |
| **Depuración de código** | Identificar y corregir errores en un fragmento | Encuentra el error en este código que calcula el promedio de una lista |
| **Revisión y retroalimentación** | Evaluar código de un estudiante con criterios dados | Revisa este programa según legibilidad, eficiencia y buenas prácticas |
| **Elaboración de material** | Producir rúbricas, instrucciones o guías de práctica | Elabora una rúbrica sobre 10 puntos para una práctica de listas |

*Nota. Las consignas definitivas se derivan del inventario de tareas obtenido en la Fase 1.*

Las consignas se ejecutan con tres modelos de lenguaje abiertos de distinto tamaño, en torno a 3, 8 y 14 mil millones de parámetros, cuantizados a 4 bits y servidos localmente mediante un entorno de ejecución de modelos abiertos, en un equipo de escritorio representativo de los laboratorios de la institución. Las especificaciones del equipo, las versiones exactas de los modelos y la configuración de generación se reportan de forma completa para garantizar la reproducibilidad, siguiendo las recomendaciones de Henderson et al. (2020). Se aplica un diseño factorial 3 × 2 × 2 de medidas repetidas, con los factores tamaño del modelo (pequeño, mediano, grande), formulación de la consigna (extensa, concisa) y longitud de la respuesta (libre, acotada mediante un límite máximo de tokens e instrucciones explícitas de brevedad). Cada combinación se ejecuta cinco veces en orden aleatorio, con un periodo de enfriamiento entre ejecuciones, lo que produce 3.000 mediciones.

La energía se mide con CodeCarbon (Courty et al., 2024), que registra el consumo de CPU, GPU y memoria durante cada ejecución. Antes de cada bloque se mide el consumo del equipo en reposo durante cinco minutos, y ese valor se descuenta para obtener la energía atribuible a la inferencia: $E_{neta} = E_{medida} - P_{reposo} \times t$, donde t es la duración de la ejecución. Las emisiones se calculan como Em (g $CO_2e$) = E_neta (kWh) × FE (g $CO_2e$/kWh), con el factor oficial de 161,6 g $CO_2e$/kWh del Sistema Nacional Interconectado para 2024 (CTFE, 2025). Además, se registran el número de tokens de entrada y de salida y la latencia, lo que permite expresar el consumo por token y comparar los resultados con otros estudios (Samsi et al., 2023; Luccioni et al., 2024).

Cada respuesta se valora con una rúbrica analítica de cuatro criterios (corrección técnica, completitud, claridad pedagógica y pertinencia para el nivel del estudiante), puntuados de 1 a 5. Dos jueces con experiencia docente en programación califican las respuestas de manera independiente y sin conocer el modelo ni la condición que las produjo. La concordancia entre jueces se estima con el coeficiente de correlación intraclase (CCI), con un mínimo aceptable de 0,75, y las discrepancias superiores a un punto se resuelven por consenso. A partir de estos datos se calcula un índice de eficiencia calidad-huella (IECH), definido como la puntuación media de calidad dividida entre las emisiones por tarea, que permite identificar las combinaciones que ofrecen la mejor calidad por unidad de impacto.

### Fase 3: estimación de la huella de los servicios en la nube

Dado que el consumo de los servicios comerciales no puede medirse directamente, su huella se estima por escenarios, siguiendo la lógica de sensibilidad recomendada ante la dispersión de las cifras disponibles (Jegham et al., 2025). Para cada tarea del inventario se calcula la huella con tres valores de energía por consulta: un escenario bajo, basado en la medición integral de la consulta mediana reportada por Elsworth et al. (2025); un escenario central y uno alto, construidos a partir de los rangos por modelo documentados por Jegham et al. (2025) para modelos comparables a los que usan los participantes. La energía se convierte en emisiones con la intensidad de carbono informada por cada fuente. La huella local, por su parte, se analiza con el rango de factores de emisión ecuatorianos de 2022 a 2024 (Tabla 3).

**Tabla 3. Parámetros de referencia para el cálculo y la comparación de emisiones**

| Parámetro | Valor o criterio | Fuente |
|---|---|---|
| **Factor de emisión SNI Ecuador 2024 (valor base)** | 0,1616 t $CO_2$-eq/MWh (161,6 g $CO_2e$/kWh) | CTFE (2025) |
| **Rango de sensibilidad local** | 0,0920 a 0,1616 t $CO_2$-eq/MWh (estadísticas 2022 a 2024) | Informes oficiales del factor de emisión del SNI |
| **Nube, escenario bajo** | 0,24 Wh y 0,03 g $CO_2e$ por consulta de texto mediana | Elsworth et al. (2025) |
| **Nube, escenarios central y alto** | Rangos por modelo según tipo y longitud de la tarea | Jegham et al. (2025) |
| **Medición local** | Energía neta por tarea medida con CodeCarbon | Courty et al. (2024) |

*Nota. SNI: Sistema Nacional Interconectado. Los valores de nube se actualizarán si se publican mediciones más recientes antes del cierre del estudio.*

La comparación entre la ejecución local y la nube se realiza para cada categoría de tarea, considerando que un modelo local pequeño solo es una alternativa válida si alcanza una calidad aceptable, definida como una puntuación media igual o superior a 4 en la rúbrica. Con el volumen de consultas registrado en la Fase 1 se extrapola la huella semanal y semestral de las asignaturas en cada escenario, y los resultados se expresan también en equivalencias comprensibles para la comunidad universitaria.

### Estrategias de uso sostenible evaluadas

Las estrategias se derivan de la literatura y se operacionalizan como condiciones del diseño factorial: elegir el modelo según la tarea en lugar de usar siempre el de mayor tamaño (Luccioni et al., 2024; UNESCO, 2025), formular consignas concisas y acotar la longitud de las respuestas (UNESCO, 2025). Una estrategia se considera eficaz cuando reduce significativamente la energía por tarea y la diferencia de calidad respecto de la condición de referencia se mantiene dentro de un margen de equivalencia de 0,5 puntos en la rúbrica. Con las estrategias que cumplan este criterio se elaboran lineamientos de uso sostenible dirigidos a los docentes de la carrera, que se validan mediante juicio de expertos.

### Análisis de datos

Los datos de la Fase 1 se analizan con estadística descriptiva (frecuencias, medias, medianas y dispersión). Para la Fase 2 se verifica la normalidad de las distribuciones con la prueba de Shapiro-Wilk y la esfericidad con la prueba de Mauchly. H1 y H2 se contrastan mediante un análisis de varianza factorial de medidas repetidas sobre la energía neta por tarea, con corrección de Greenhouse-Geisser cuando corresponda; si no se cumplen los supuestos, se aplica una transformación logarítmica o, en su defecto, pruebas no paramétricas de Friedman y de rangos con signo de Wilcoxon con corrección de Holm. La equivalencia de calidad se evalúa con el procedimiento de dos pruebas unilaterales (TOST) con el margen definido. Se reportan tamaños del efecto ($\eta^2$ parcial y r) e intervalos de confianza del 95 %, con un nivel de significación de $\alpha = 0,05$. Para H3, la comparación entre la huella local y la estimada para la nube se presenta por categoría de tarea y escenario, y la incertidumbre se examina con una simulación de Monte Carlo que combina los rangos de los factores de emisión y de energía por consulta. El análisis se realiza en Python y los scripts se publican en un repositorio abierto.

## Consideraciones éticas

La participación es voluntaria y se formaliza mediante consentimiento informado, en el que se explican el propósito del estudio, el tipo de información registrada y el derecho a retirarse sin consecuencias académicas. El registro de uso no recoge el contenido de las conversaciones con las herramientas de IA, y los datos se anonimizan con códigos numéricos, conforme a la Ley Orgánica de Protección de Datos Personales del Ecuador (Asamblea Nacional del Ecuador, 2021). La información no se utiliza con fines de calificación. El estudio cuenta con la autorización de las autoridades de la carrera antes del inicio de la recolección de datos.

## Referencias

Asamblea Nacional del Ecuador. (2021). *Ley Orgánica de Protección de Datos Personales* (Registro Oficial Suplemento 459, 26 de mayo de 2021).

Becker, B. A., Denny, P., Finnie-Ansley, J., Luxton-Reilly, A., Prather, J., & Santos, E. A. (2023). Programming is hard – or at least it used to be: Educational opportunities and challenges of AI code generation. En *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (pp. 500–506). ACM. https://doi.org/10.1145/3545945.3569759

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? En *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Comisión Técnica de determinación de Factores de Emisión de Gases de Efecto Invernadero. (2025). *Factor de emisión de $CO_2$ del Sistema Nacional Interconectado de Ecuador: Informe 2024*. Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables. https://arconel.gob.ec/wp-content/uploads/downloads/2025/09/Factor-de-emision-de-CO2-del-Sistema-Nacional-Interconectado-de-Ecuador-Informe-2024_compressed.pdf

Courty, B., Schmidt, V., Luccioni, S., Goyal-Kamal, MarionCoutarel, Feld, B., Lecourt, J., LiamConnell, Saboni, A., Inimaz, supatomic, Léval, M., Blanche, L., Cruveiller, A., ouminasara, Zhao, F., Joshi, A., Bogroff, A., de Lavoreille, H., … MinervaBooks. (2024). *mlco2/codecarbon: v2.4.1* [Software]. Zenodo. https://doi.org/10.5281/zenodo.11171501

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3.ª ed.). SAGE.

de Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule, 7*(10), 2191–2194. https://doi.org/10.1016/j.joule.2023.09.004

Denny, P., Prather, J., Becker, B. A., Finnie-Ansley, J., Hellas, A., Leinonen, J., Luxton-Reilly, A., Reeves, B. N., Santos, E. A., & Sarsa, S. (2024). Computing education in the era of generative AI. *Communications of the ACM, 67*(2), 56–67. https://doi.org/10.1145/3624720

Elsworth, C., Huang, K., Patterson, D., Schneider, I., Sedivy, R., Goodman, S., Townsend, B., Ranganathan, P., Dean, J., Vahdat, A., Gomes, B., & Manyika, J. (2025). *Measuring the*

*environmental impact of delivering AI at Google scale* (arXiv:2508.15734). arXiv. https://doi.org/10.48550/arXiv.2508.15734

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research, 21*(248), 1–43. https://jmlr.org/papers/v21/20-312.html

Hernández-Sampieri, R., & Mendoza, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education.

International Energy Agency. (2025). *Energy and AI*. IEA. https://www.iea.org/reports/energy-and-ai

Jegham, N., Abdelatti, M., Koh, C. Y., Elmoubarki, L., & Hendawi, A. (2025). *How hungry is AI? Benchmarking energy, water, and carbon footprint of LLM inference* (arXiv:2505.09598). arXiv. https://doi.org/10.48550/arXiv.2505.09598

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Artículo 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lannelongue, L., Grealey, J., & Inouye, M. (2021). Green algorithms: Quantifying the carbon footprint of computation. *Advanced Science, 8*(12), Artículo 2100707. https://doi.org/10.1002/advs.202100707

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). *Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models* (arXiv:2304.03271). arXiv. https://doi.org/10.48550/arXiv.2304.03271

Luccioni, A. S., Jernite, Y., & Strubell, E. (2024). Power hungry processing: Watts driving the cost of AI deployment? En *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 85–99). ACM. https://doi.org/10.1145/3630106.3658542

Luccioni, A. S., Strubell, E., & Crawford, K. (2025). From efficiency gains to rebound effects: The problem of Jevons' paradox in AI's polarized environmental debate. En *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* (pp. 76–88). ACM. https://doi.org/10.1145/3715275.3732007

Luccioni, A. S., Viguier, S., & Ligozat, A.-L. (2023). Estimating the carbon footprint of BLOOM, a 176B parameter language model. *Journal of Machine Learning Research, 24*(253), 1–15. https://jmlr.org/papers/v24/23-0069.html

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Naciones Unidas. (2015). *Transformar nuestro mundo: La Agenda 2030 para el Desarrollo Sostenible* (A/RES/70/1). https://sdgs.un.org/2030agenda

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). *Carbon emissions and large neural network training* (arXiv:2104.10350). arXiv. https://doi.org/10.48550/arXiv.2104.10350

Samsi, S., Zhao, D., McDonald, J., Li, B., Michaleas, A., Jones, M., Bergeron, W., Kepner, J., Tiwari, D., & Gadepally, V. (2023). From words to watts: Benchmarking the energy costs of large language model inference. En *2023 IEEE High Performance Extreme Computing Conference (HPEC)* (pp. 1–9). IEEE. https://doi.org/10.1109/HPEC58863.2023.10363447

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM, 63*(12), 54–63. https://doi.org/10.1145/3381831

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. En *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645–3650). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1355

UNESCO. (2022). *Recomendación sobre la ética de la inteligencia artificial*. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

UNESCO. (2025). *Smarter, smaller, stronger: Resource-efficient generative AI & the future of digital transformation*. https://www.unesco.org/en/articles/ai-large-language-models-new-report-shows-small-changes-can-reduce-energy-use-90

Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. *WIREs Data Mining and Knowledge Discovery, 13*(4), Artículo e1507. https://doi.org/10.1002/widm.1507

Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Behram, F. A., Huang, J., Bai, C., Gschwind, M., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H.-H. S., … Hazelwood, K. (2022). Sustainable AI: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems, 4*, 795–813. https://proceedings.mlsys.org/paper_files/paper/2022/hash/462211f67c7d858f663355eff93b745e-Abstract.html