VISUAL REVIEW | Vol. 18, No. 6, 2026 | ISSN 2695-9631
International Visual Culture Review / Revista Internacional de Cultura Visual
https://doi.org/10.62161/revvisual.v18.6452

MÁS ALLÁ DE LA NUBE
Ética, descentralización, ecología digital y gobernanza algorítmica de la Inteligencia Artificial local

JOAN HERNANDEZ-SERRET [1]
jhernandezs@uic.es
NURIA ROCA-TRENCHS [1]
nroca@uic.es
LARA MARTIN-VICARIO [1]
lmartinv@uic.es
[1]Universitat Internacional de Catalunya, Spain

PALABRAS CLAVE

gobernanza algorítmica
inteligencia artificial local
poder digital
ética de la IA,
sostenibilidad digital
descentralización tecnológica

RESUMEN

Se analiza cómo los modelos de IA visual ejecutados en dispositivos domésticos reconfiguran la gobernanza algorítmica, la seguridad y la distribución del poder. Desde la democratización tecnológica, el estudio se centra en ordenadores de uso cotidiano, sin GPU internas ni externas de alta gama, para evaluar una IA energéticamente eficiente, éticamente gobernable y menos dependiente de plataformas corporativas. Prescindir de GPU evita renovar equipos y reduce el impacto ambiental. Se adopta una metodología comparativa cualitativa, basada en experimentos controlados con modelos de IA visual locales y en la nube, aplicados a diversas tipologías de imágenes y con variaciones sistemáticas de parámetros algorítmicos. Los resultados muestran que la IA visual local, aunque de rendimiento limitado, resulta suficiente para tareas descriptivas. Su uso, gobernado mediante control algorítmico, puede redistribuir el poder tecnológico y ofrecer alternativas a la centralización en la nube, reforzando la autonomía, la privacidad y la capacidad institucional.

Recibido: 03/ 07 / 2026
Aceptado: 05/ 10 / 2026

1. Introducción

Aunque los sistemas de inteligencia artificial (IA) se han desplegado de forma muy rápida, su desarrollo se ha centrado mayoritariamente en modelos a gran escala ejecutados en la nube, priorizando métricas de rendimiento y escalabilidad por encima del control, la sostenibilidad, la circulación de contenidos o la gobernanza (OECD, 2022; Strubell et al., 2019). Este proceso ha contribuido a un aumento de la centralización computacional, con elevados costes energéticos y una creciente concentración de poder en manos de grandes plataformas tecnológicas que condicionan el acceso, el control y la circulación de contenidos (Crawford, 2021). Autores como Schwarts et al. (2019) han cuestionado, además, el impacto ambiental de estos modelos a gran escala y su dependencia de infraestructuras energéticamente intensivas.

Los sistemas de IA ejecutados en local deben ser entendidos como modelos de IA generativa descargados y ejecutados desde dispositivos domésticos, a los que se referirá como IA local de ahora en adelante. Estos modelos, frente a las modelos en nube, no sólo garantizan la privacidad de los datos, sino que redistribuyen el poder algorítmico, permiten una gobernanza ética situada, refuerzan la autonomía institucional y reducen tanto la dependencia corporativa como el impacto ecológico. Wang et al. (2025) señalan una serie de ventajas de estos modelos respecto a los ejecutados en la nube. Por un lado, resultan más sostenibles, al disminuir el uso de centros de datos y la transferencia masiva y redundante de información. Por el otro, al permanecer bajo control directo del usuario, se asegura la privacidad y la soberanía de los datos. Por ejemplo, en el caso de la IA local para el análisis de imágenes, estas últimas no salen de los dispositivos por lo que no existe transferencia a terceros ni almacenamiento en servidores corporativos. Ello facilita el cumplimiento normativo en materia de protección de datos -incluyendo el Reglamento General de Protección de Datos (RGPD)- protección de menores y entornos sensibles. Por el otro, se produce una redistribución del poder tecnológico, ya que el control deja de estar en manos de plataformas centralizadas para situarse en decisiones locales sobre qué modelo utilizar, cómo configurarlo y cuándo aplicarlo. Esta capacidad habilita una gobernanza algorítmica efectiva, basada en el ajuste de parámetros como la temperatura o el top-p, donde la ética deja de ser una política corporativa para convertirse en una decisión técnica local, incluida la aplicación de prompts éticos situados. Finalmente, el uso de IA visual local fortalece las autonomías institucionales, al no requerir licencias, permitir su uso sin conexión y no estar sujeta a cambios en las políticas de las plataformas.

En este contexto, los modelos de IA local han sido relativamente poco explorados y rara vez se han considerado como una alternativa técnica, política y ecológica a la centralización computacional en la nube, en gran medida debido a la aceleración del cambio tecnológico, que ha dificultado un análisis sistemático y situado de sus potencialidades y límites. Desde una lectura sociotécnica, esta invisibilización es coherente con lo que Latour (1993) describe como el efecto de las infraestructuras dominantes: cuanto más funcionan y se expanden, más opacas se vuelven y menos visibles resultan sus mediaciones materiales y políticas. Según Latour, la modernidad se sostiene mediante dos conjuntos de prácticas simultáneas: la purificación, que produce dos zonas ontológicas diferenciadas: la de los no humanos (naturaleza) y la de los humanos (sociedad/cultura); y la mediación o traducción, que genera híbridos que mezclan lo social y lo material (Latour, 1993). Sin embargo, la llamada Constitución moderna oculta sistemáticamente el trabajo de mediación, mientras hace visible y legítimo únicamente el trabajo de purificación (Latour, 1993). A su vez, en la línea de Yuk Hui (2016), la aceleración tecnológica tiende a consolidar cosmologías técnicas hegemónicas, dificultando la articulación de alternativas locales donde los regímenes de energía, sentido y gobernanza puedan definirse en proximidad.

En este trabajo, se concibe la IA como un sistema integral de relaciones, flujos e impactos que emergen cuando la producción, el procesamiento y la gobernanza de imágenes se reorganizan fuera de infraestructuras centralizadas, atendiendo simultáneamente a dimensiones técnicas, políticas, energéticas y éticas. Se plantea que la IA local aplicada al tratamiento de la imagen, aun ofreciendo un rendimiento limitado pero suficiente para determinadas tareas descriptivas, puede redistribuir el poder tecnológico, reducir el impacto ecológico y ofrecer alternativas éticas y seguras a la centralización en la nube, reforzando la autonomía, la privacidad y la gobernanza mediante el control algorítmico local.

A partir de este marco, la investigación se centra en un análisis comparativo del uso de IA visual en la nube y en local aplicado a la creación y descripción de imágenes, con el objetivo de examinar sus

implicaciones en términos de poder, gobernanza algorítmica, sostenibilidad y control tecnológico. Por IA visual se entenderá aquellos modelos de IA generativa que tienen la capacidad multimodal de analizar e interpretar imágenes. Por ello, el presente trabajo plantea cómo difiere el comportamiento descriptivo e interpretativo de los modelos de IA visual ejecutados localmente en comparación con los modelos basados en la nube. Esta comparativa se fundamenta en términos de precisión, estabilidad, alucinaciones y sesgos, y en qué tipos de tareas y contextos de uso la IA local resulta suficiente sin requerir infraestructura en la nube, especialmente en contextos sensibles. Los contextos sensibles serían aquellos en los que una interacción o acción mínima pueden producir cambios significativos en términos de resultado (Yau y Karim, 2004). Estos cambios podrían ser irreversibles o costos en términos de resultado.

Es por ello que la presente investigación tiene como objetivo principal analizar el comportamiento de modelos de IA aplicados a la imagen ejecutados en local y en la nube en tareas de descripción e interpretación de imágenes. De esta manera, se pretende identificar en qué tipos de tareas y contextos de uso la IA local resulta funcionalmente suficiente y en cuáles es necesaria la infraestructura en la nube.

Como novedad, esta investigación evalúa de forma integrada cómo la ejecución local incide en el comportamiento descriptivo e interpretativo de los modelos, en la transparencia y trazabilidad del control paramétrico, en la suficiencia funcional para usos cotidianos, en la reducción de la huella energética y en la reconfiguración de las relaciones de dependencia tecnológica, un enfoque abordado hasta ahora de forma incipiente (Wang et al., 2025; Zhang et al., 2022). Este planteamiento transversal permite un examen crítico de cómo la IA visual local puede contribuir a una gobernanza algorítmica más democrática, a una mayor sostenibilidad y a una autonomía efectiva del usuario.

2. Marco teórico

Según su arquitectura, control y repercusiones sociales, se distingue entre la IA en la nube y la IA local. La primera centraliza el procesamiento y almacenamiento de datos en infraestructuras externas que son gestionadas por grandes plataformas, lo que potencia el aprendizaje a gran escala y el control de un gran volumen de información, por lo que la autonomía de los usuarios es limitada. La segunda opera directamente en dispositivos del usuario o en entornos controlados por comunidades, lo que permite una mayor privacidad de los datos. Su potencial de aprendizaje y predicción puede ser menor que en la nube, pero el usuario sigue siendo autónomo y se reduce la dependencia de plataformas centralizadas, respetando los derechos individuales (Jane, 2025).

La IA visual se ha desarrollado, sobre todo, en la nube, dónde los modelos de gran escala son entrenados y ejecutados en infraestructuras centralizadas y de alto rendimiento. Esto implica una gran dependencia de centros de datos energéticamente intensivos, una transparencia operativa muy limitada y una concentración del control técnico y económico en grandes plataformas (Crawford, 2021; OECD, 2022; Strubell et al., 2019). Los modelos de IA visual local que son ejecutados en dispositivos domésticos se plantean como una alternativa que procesa los datos mejorando la privacidad a la vez que disminuyen la huella energética y habilitan formas más gobernables de producción y circulación de contenidos porque operan con restricciones de hardware y a menor escala (Moosman, 2025; Schwartz et al, 2020).

2.1. Gobernanza, control y poder

La IA ha sido analizada por distintos autores como dispositivo que redefine las formas de ejercicio de poder, la gobernanza y la ética en la sociedad contemporánea. Por ejemplo, Zuboff (2019) considera la IA como un instrumento del capitalismo de vigilancia, basado en la extracción y la valoración predictiva de un gran volumen de datos, que puede generar asimetrías de poder. En la misma línea, Couldry y Mejías (2019) describen este proceso como colonialismo de datos y destacan como la concentración de infraestructuras y plataformas refuerzan desigualdades estructurales existentes. Rouvroy y Berns (2018) señalan que la gubernamentalidad algorítmica que mediante el tratamiento automatizado de datos desactiva el sujeto político y alertan de la substitución de la decisión colectiva por la optimización técnica. Mientras, Pascuale (2015) señala la opacidad de los algoritmos y su capacidad de concentrar poder y control de empresas y sectores clave (p.ej.: la economía, la información o las plataformas digitales) lo que refuerza la necesidad de mecanismos públicos de transparencia para evitar abusos de

poder. Es por todo ello que Bratton (2016) propone entender la IA no sólo como una tecnología sino como un elemento crítico de las estructuras de poder.

La distinción entre IA local y IA en la nube no solo es técnica, sino también política, ya que implica distintos regímenes de gobernanza, control y distribución del poder. Desde esta perspectiva, y en diálogo con Foucault (1975), la IA no sería solo una herramienta técnica, sino que se puede entender como una tecnología de poder que organiza, clasifica y produce conocimiento sobre los individuos y las sociedades, en línea con el concepto del panóptico y de las lógicas de poder disciplinario. En este sentido, la IA no se limitaría a reprimir o excluir, sino que tiene la capacidad de generar nuevas realidades, objetos y verdades. Así, los algoritmos de IA pueden influir en la percepción de la realidad mediante la selección de información, la creación de contenido y la toma de decisiones automatizadas. Esto implica una reconfiguración de las relaciones de poder, ya que la IA puede operar de manera autónoma y perpetuar dinámicas de control sin necesidad de intervención humana directa. Esto plantea cuestiones éticas fundamentales ya que la implementación y uso de la IA puede reforzar desigualdades, perpetuar sesgos y afectar la autonomía individual, generando vulnerabilidades sociales derivadas del uso de la inteligencia artificial (Benjamins y Salazar, 2020; Blanco et al., 2019).

La descentralización y la soberanía tecnológica cuestionan el modelo de inteligencia artificial basado en grandes plataformas centralizadas y servicios en la nube. Para Morozov (2013), el solucionismo tecnológico refuerza la dependencia de infraestructuras privadas y, a menudo, opacas, por lo que los gobiernos y la ciudadanía ven limitada la capacidad de decidir sobre el uso de datos. Esto se traduce en una concentración de poder computacional, económico y político en muy pocas empresas y dificulta el desarrollo de alternativas locales o públicas. Mazzucato (2021) propone formas de organización más distribuidas y orientadas al interés público y Bauwens (2005) propone la creación de comunidades digitales y una gobernanza distribuida para reducir la dependencia de la nube.

2.2. Tecnología responsable y ética situada

Los usos de la IA, tanto en la nube como en local, se entienden como una práctica socio-técnica en la que sus efectos dependen del contexto en el que se diseña, implementa y gobierna. Floridi (2019), introduce los principios para una gobernanza responsable de la información, lo que implica introducir criterios éticos en la regulación e implementación de tecnologías para proteger la autonomía de las personas. En la misma línea, Eubanks (2020) alerta sobre las desigualdades sociales que potencian tecnologías como la IA, limitando su acceso a derechos y recursos, por lo que concluye, también, en la necesidad de regular con un enfoque de justicia social. Todo ello, coincide con lo que señala Nissenbaum (2010), en su idea de privacidad contextual, considera que un uso adecuado de la información depende del contexto social y normativo en el que se produce, así, el diseño de tecnologías respetuosas debería tener en cuenta la privacidad como el cumplimiento de normas sociales que regulen la circulación de información. Y Jasanoff (2016) señala que la innovación tecnológica no es neutra y que su impacto depende de cómo se diseña, regula y usa en contextos sociales específicos, por lo que propone una responsabilidad compartida entre todos los agentes que participan en la gobernanza tecnológica.

2.3. Ecología computacional y sostenibilidad

Los sistemas de inteligencia artificial son infraestructuras materiales con gran impacto social y ambiental. Crawford (2021) subraya cómo su funcionamiento depende de la extracción de minerales, de un consumo energético elevado y de la generación de un gran volumen de residuos electrónicos. A ello, Ensmenger (2021) añade los costos sociales de su uso, como el impacto laboral de mantener centros de datos (a menudo con trabajo precarizado e invisibilizado) y de procesamiento masivo de datos. Por ello, el uso de la IA se convierte, también, en una cuestión ecológica y centra el debate sobre sostenibilidad por el impacto de su huella ecológica y social. Todo ello toma sentido en un contexto global en la que la computación contemporánea se entiende, tal y cómo sostiene Bratton (2015), como una arquitectura planetaria compuesta por múltiples capas interrelacionadas: la infraestructura física (energía, centros de datos, cables, satélites), las plataformas, los usuarios y las formas de gobernanza. La computación se apoya en sistemas energéticos, extractivos y logísticos a escala mundial, por lo que sus consecuencias a nivel ecológico son estructurales. De hecho, los centros de datos consumieron 415 TWh de electricidad en 2024, esto es un centro de datos grande puede consumir la electricidad de 100.000 hogares en un año. Y, aunque solo representa el 1,5% de la electricidad global el 2024, la demanda de energía crece rápido y se espera que se duplique en 2030 hasta los 945 TWh (European

Parliamentary Research Service, 2025). En cambio si se opta por el uso de modelos más pequeños (como los locales), se podría reducir el consumo energético global en un 27,8%, unos 31,9% en 2025 (da Silva Barros et al., 2025).

3. Metodología

La investigación adopta una metodología comparativa cualitativa basada en experimentación controlada, estructurada mediante comparaciones por pares entre modelos de IA visual ejecutados en entornos locales domésticos y modelos equivalentes basados en la nube. Para asegurar la comparabilidad de los resultados, se aplicaron las mismas tareas, *prompts* y tipologías de imágenes en ambos entornos, variando únicamente las condiciones de ejecución, la infraestructura computacional y el grado de control paramétrico. Este enfoque permite analizar el comportamiento de los modelos en condiciones funcionalmente distintas, aislando el impacto de la centralización, el control algorítmico y la dependencia de plataforma sobre los resultados.

Se ha procedido a un análisis sobre un conjunto homogéneo de imágenes, organizado en seis tipologías: escenas urbanas, entornos naturales, imágenes nocturnas, naturaleza inanimada (bodegones), composiciones abstractas (*light painting*) y escenas deportivas dinámicas, tomando como referencia las definiciones de tipologías de la Federació Catalana de Fotografia (2023). Cada imagen ha sido procesada por todos los modelos siguiendo un protocolo de prompts idéntico, compuesto por tres tareas secuenciales:

- Prompt para la descripción visual factual: '*Describe with extreme precision all visible objects in the image, including colors, shapes, and spatial relationships. Do not invent anything*'.
- Prompt para la interpretación de la escena y de la acción: '*Explain what is happening in the image, what action the people or objects are performing, and what elements allow you to deduce it'*.
- Prompt para la síntesis breve y generación de títulos: '*Summarize the main content of the image in three clear sentences. Then generate three alternative titles without adding unnecessary details*'.

Para el análisis del entorno local se emplearon modelos de IA visual multimodal de tamaño reducido y medio, ejecutados en dispositivos domésticos de gama media mediante plataformas locales, concretamente: Qwen2.5-VL-7B-Instruct, con parámetros estándares (Top P 0,9, Top k 40, Temperatura 0,6) y Qwen3-VL-2B-Instruct con parámetros restrictivos (Top-p 0.5; Top-k 15; Temperatura 0.05). Estos modelos fueron seleccionados por su viabilidad en hardware doméstico sin GPUs de alta gama. En el entorno en la nube se utilizaron modelos de mayor escala y capacidad visual, específicamente: Qwen3-VL-32B y Qwen3-VL-235B-A22B. En los modelos locales se introdujeron variaciones sistemáticas de parámetros algorítmicos (temperatura, top-p, top-k) con el objetivo de observar su impacto en estabilidad, alucinaciones, sesgos y control interpretativo. Los modelos en la nube se analizaron con sus configuraciones por defecto, reflejando las condiciones reales de uso en plataformas centralizadas.

Los resultados se evaluaron comparativamente atendiendo a parámetros de: precisión descriptiva; estabilidad de salida (p.ej.: la aparición de alucinaciones, el control semántico y trazabilidad del proceso); adecuación a contextos variables y sensibles (imágenes distintas, con distintas características, como pudiera ser la representación de escenas dinámicas). También se han aplicado criterios de conveniencia funcional (puede hacer lo que el usuario le pida), energética (no depende de los recursos externos como la nube) y de gobernanza (el usuario lo puede controlar). Por precisión descriptiva se entenderá como la capacidad del modelo para describir lo que realmente aparece en la imagen (p.ej.: identificación correcta de los elementos visibles, características y relaciones) sin añadir información que no está presente. Por su parte, la gobernanza algorítmica se define aquí como la capacidad de controlar el modelo a nivel local, incluso sin conexión a red, garantizando autonomía, soberanía tecnológica y seguridad mediante la no externalización de datos a la nube.

Para realizar el análisis el primer y el segundo autor emplearon las mismas características técnicas tanto en el uso de equipo informático, modelos de IA (local y en nube), prompts e imágenes empleadas. Una vez devuelto el análisis de la IA, se compararon y contrastaron sistemáticamente los resultados obtenidos en los experimentos empleados. Este diseño permite identificar en qué tipos de tareas la IA local resulta suficiente y cuándo el uso de modelos en la nube aporta ventajas significativas, aportando evidencias para una toma de decisiones informada sobre configuraciones locales, en la nube o híbridas.

4. Análisis

Esta investigación presenta un análisis comparativo de la capacidad de visión computacional entre modelos locales (Qwen3-VL-2B) de IA y modelos de IA en la nube (Qwen3-VL-235B y Qwen3-MAX), utilizando un *dataset* diversificado de seis tipologías de imagen: (1) arquitectura urbana (figura 1); (2) naturaleza (figura 2); (3) nocturna (figura 3); (4) bodegón (figura 4); (5) arte abstracto (figura 5); y acción deportiva (figura 6), según la clasificación de la Federació Catalana de Fotografia (2023).

Figura 1. Fotografía empleada para la tipología arquitectura urbana.



Fuente: https://depositphotos.com/nito103, 2011

Figura 2. Fotografía empleada para la tipología naturaleza.



Fuente: iStock/FG Trade, 2018

Figura 3. Fotografía empleada para la tipología nocturna.



Fuente: iStock/Florian Parzefall, 2020

Figura 4. Fotografía empleada para la tipología bodegón.



Fuente: iStock/RCerruti, 2012

Figura 5. Imagen empleada para la tipología arte abstracto.



Fuente: iStock/Wacomka, 2022

Imagen 6. Fotografía empleada para tipología acción deportiva.



Fuente: iStock/Simonkr, 2017

Los experimentos devolvieron una gran cantidad de datos que, por cuestiones de extensión del texto y claridad en la exposición, se expondrán únicamente aquellos resultados destacables desde el punto de vista de eficacia y eficiencia en relación a la descripción y la síntesis breve y la generación de títulos de las imágenes. Es por ello que, los resultados se estructuran en tres ejes: (1) precisión descriptiva visual; (2) rendimiento en la síntesis de información; (3) rendimiento según la complejidad semántica y dinámica de las imágenes.

4.1. Precisión descriptiva visual

Se observa jerarquía en la "mirada" algorítmica. Los modelos locales operan, sobre todo, bajo una percepción general y superficial, mientras que los modelos en la nube alcanzan una percepción estructural o experta. Los modelos locales identifican correctamente categorías básicas, pero no logran captar la estructura volumétrica ni los detalles técnicos de la escena presentada. Por ello, tienden a recurrir a la sustitución de rasgos alucinados, completando los vacíos perceptivos con prototipos internos. Esto se traduce en errores críticos, como la identificación incorrecta del Arco de Triunfo (figura 1) como "Arc de la Mercè" o la invención de elementos arquitectónicos inexistentes. En cambio, los modelos en la nube demuestran una competencia visual estructural muy superior, percibiendo materiales (madera envejecida, metal oxidado), geometrías arquitectónicas precisas y procesos naturales complejos, como la erosión en ríos. En imágenes de alta complejidad visual, por ejemplo, light painting (figura 5), son los únicos capaces de reconstruir correctamente la geometría tridimensional de los trazos de luz.

Cuando se aplican parámetros restrictivos (top-p 0.5, top-k 15, T=0.05) se observa que estos actúan como un mecanismo de contención: reducen las alucinaciones y mejoran la literalidad, pero no incrementan la capacidad perceptiva real del modelo. Con ellos, el modelo local deja de inventar, pero sigue sin "ver" el detalle fino. Por ejemplo, en la figura 4, el modelo local pasa de describir "Los elementos arquitectónicos (arcos abovedados de ladrillo, soportes de piedra o ladrillo para los barriles y la disposición general) sugieren una bodega histórica o de estilo tradicional [...]." a una salida más literal como "La imagen muestra una bodega subterránea con un largo pasillo bordeado de grandes barriles de madera. El espacio está construido con paredes de ladrillo arqueadas y suelo de hormigón." En cambio, el modelo MAX realiza una lectura arquitectónica profunda de la bodega, detectando la textura de la madera envejecida, el óxido en los metales y la simetría rítmica de los arcos de ladrillo. En el ejemplo del Arco de Triunfo, el modelo local ve un monumento, pero a menudo inventa arcos laterales o se equivoca en el nombre. En cambio, los modelos centralizados perciben con exactitud los frisos, las esculturas y la profundidad del conjunto urbano.

Así, el uso de la IA en la nube implica que la interpretación de la imagen depende de servidores corporativos que apliquen filtros de moderación y políticas de censura sobre las que el usuario no tiene control. Sin embargo, un modelo local permite al usuario ajustar y controlar parámetros de salida como

la temperatura o el top-p para obtener una descripción estrictamente literal, garantizando la privacidad y la autonomía del relato, aunque la capacidad visual sea menor.

4.2. Rendimiento en tareas de resumen

Un hallazgo destacable es que la tarea de resumen no resulta discriminante para evaluar la capacidad de análisis visual de los modelos, ya que se obtiene el mismo resultado. En las seis imágenes analizadas, las diferencias en la descripción extrema desaparecen en la fase de síntesis. Todos los modelos logran producir resúmenes funcionalmente equivalentes, lo que igualan o enmascara las deficiencias de visión de los modelos más pequeños.

En tareas de identificación de categorías y entidades básicas, los modelos locales muestran un rendimiento comparable al de los modelos en la nube. En imágenes estáticas y de baja complejidad semántica como el bodegón o arquitectura, ambos identifican correctamente cuerpos animados e inanimados y estáticos, como personas, animales y objetos, así como la naturaleza general del entorno (espacios naturales, urbanos o interiores). En cambio, en imágenes nocturnas analizadas en local, que implican baja iluminación, aumenta la ambigüedad, se produce una pérdida del detalle y existen más alucinaciones.

En segundo lugar, las tareas de resumen e indexación, que constituyen lo que puede denominarse una "zona de neutralización", no permiten diferenciar la calidad visual de los modelos. Cuando el objetivo es generar una breve descripción informativa (por ejemplo, un resumen de tres frases o un título descriptivo), el modelo local resulta tan útil como el modelo en la nube. Aunque este último suele aportar matices estilísticos o terminológicos más refinados, la función comunicativa básica se cumple en ambos casos.

Asimismo, en escenas estáticas de naturaleza (figura 2) o imágenes nocturnas sin acción compleja (figura 3), la brecha de rendimiento es limitada. El modelo local identifica correctamente elementos fundamentales como ríos, vegetación, cielo, siluetas o reflejos, lo que permite construir descripciones atmosféricas básicas, aunque no es capaz de reconocer procesos geomorfológicos o dinámicas lumínicas avanzadas, su desempeño resulta suficiente para usos descriptivos generales. En este contexto, los modelos locales también muestran una capacidad adecuada para describir paletas cromáticas e iluminación global, facilitando una clasificación estética elemental de la imagen.

Por el contrario, los modelos en la nube presentan una capacidad visual estructural significativamente superior, que les permite captar detalles técnicos y arquitectónicos con alta precisión. Esta densidad informativa puede derivar en interpretaciones sobreanalíticas, en las que se enfatizan texturas, simetrías o procesos físicos que no siempre son necesarios para una descripción simple. Mientras que los modelos locales se limitan a una percepción general, los modelos densos analizan con profundidad la estructura espacial, la física de la luz o los materiales, lo que puede generar una sobrecarga informativa para usuarios que buscan únicamente una descripción básica.

Estos resultados ponen de manifiesto que, aunque los modelos en la nube "perciben" mucho más que los modelos locales, dicha ventaja se neutraliza en tareas de síntesis breve, donde la granularidad visual no es un requisito. En consecuencia, los resúmenes generados por ambos tipos de modelos tienden a ser funcionalmente equivalentes.

La suficiencia del modelo local se ve superada cuando la tarea exige precisión nominal, interpretación técnica o táctica (como en escenas deportivas) o comprensión volumétrica y espacial. En estos casos, los modelos locales pierden fiabilidad, mientras que los modelos en la nube mantienen una interpretación visual robusta.

Así, los modelos locales resultan adecuados para tareas de etiquetado general, resumen, indexación y contextos donde la gobernanza, la privacidad y la autonomía del usuario son prioritarias. Sin embargo, para tareas que requieren una veracidad visual fuerte o un análisis experto, la dependencia de modelos en la nube sigue siendo necesaria. Las fuentes analizadas coinciden en que la superioridad perceptiva de estos últimos es evidente en tareas de percepción estructurada, aunque a menudo quede oculta o resulte innecesaria en tareas textuales superficiales.

4.3. Complejidad semántica y dinámica de la imagen

Los resultados muestran que la magnitud de la diferencia entre los modelos locales y los modelos en la nube no es uniforme, sino que varía de forma significativa en función de la complejidad semántica y dinámica del contenido visual analizado. La arquitectura del modelo incide directamente en su

competencia visual estructural, generando una brecha de rendimiento que se amplía según la naturaleza de la escena.

Por ejemplo, en escenas estáticas o naturales, como paisajes diurnos, ríos sinuosos o imágenes nocturnas sin acción compleja, la diferencia entre ambos tipos de modelos es moderada. En estos contextos, los modelos locales resultan suficientes para una identificación básica de los elementos presentes en la imagen. Su desempeño se apoya en la identificación de categorías globales (como agua, vegetación, cielo o siluetas humanas), lo que permite construir descripciones funcionales, aunque carentes de profundidad estructural.

Esto se explica por el uso de mecanismos de identificación por prototipos, también descritos como *hallucinated feature substitution* (sustitución por característica alucinada). En escenas de baja complejidad semántica, el modelo local reconoce patrones generales y completa la descripción mediante representaciones internas estandarizadas. Dado que el criterio de corrección es eminentemente categórico (por ejemplo, identificar si se trata de un río o de una escena nocturna), los errores perceptivos resultan menos evidentes para el usuario. Además, en este tipo de imágenes no existe una "regla" o "táctica" que interpretar: la tarea es fundamentalmente lingüística y atmosférica, lo que permite al modelo local describir gradientes cromáticos, iluminación general o disposición de elementos sin necesidad de comprender procesos geomorfológicos o estructuras fotográficas profundas, capacidades que sí exhiben los modelos más densos. Contrariamente, en escenas dinámicas y socialmente estructuradas, como las acciones deportivas, la brecha entre modelos locales y de la nube alcanza su máxima expresión. El caso del fútbol (figura 6) constituye un ejemplo paradigmático de tarea perceptiva estructurada, ya que la correcta interpretación de la imagen exige una reconstrucción espacial, cinemática y táctica que va mucho más allá de la simple identificación de objetos.

En este contexto, se observa que los modelos en la nube (235B y MAX) actúan de forma análoga a un analista profesional de VAR. Analizan con precisión la orientación corporal de los jugadores, las líneas de fuga, la trayectoria del balón y la relación espacial entre los distintos elementos de la escena. Esta capacidad les permite determinar si el balón ha cruzado la línea de gol o si el portero ha llegado tarde a la intervención, inferencias que requieren una representación multimodal de alta resolución. En cambio, el modelo local se limita a descripciones planas y genéricas, como "un jugador chuta el balón", sin identificar acciones técnicas específicas ni interpretar la situación táctica. Esta limitación conduce a errores de interpretación crítica, tales como afirmar que el portero ha detenido el balón cuando la imagen muestra que ha sido superado, o introducir jugadas estáticas inexistentes (como un "tiro libre") en medio de una acción dinámica.

5. Discusión y conclusiones

El presente estudio tenía como objetivo identificar en qué tipos de tareas y contextos de uso la IA local resulta funcionalmente suficiente y en cuáles es necesaria la infraestructura en la nube. Los resultados ponen de relieve que las diferencias observables entre modelos de IA visual ejecutados localmente y modelos centralizados en la nube no pueden interpretarse solamente según su rendimiento técnico, sino que deben entenderse como asimetrías epistémicas: qué tipo de "verdad visual" (es decir, aquello que se asume por la mayoría como evidente, certero y contrastable) se produce y con qué fiabilidad; y de gobernanza algorítmica (quién controla la infraestructura, los parámetros, los datos y , en última instancia, el relato). Además, dichas diferencias tienen implicaciones políticas, ecológicas y sociales (Floridi, 2019; Jasanoff, 2016).

El primer hallazgo observado es que los modelos locales presentan una suficiencia funcional condicionada. Por un lado, son capaces de realizar tareas descriptivas básicas -como el etiquetado general, la clasificación de escenas o la generación de resúmenes breves-. Por ejemplo, se observan alucinaciones nominales y estructurales como la confusión de Arc del Triomf por Arc de la Mercè, es decir, no tiene capacidad para identificar estructuras volumétricas, ni tienen precisión nominal o dinámica espacial, por lo que deviene vulnerable a errores y al fenómeno de sustitución de características alucinadas. Por ello, se deduce que la IA local resulta adecuada o suficiente para contextos de baja complejidad semántica y dinámica, mientras que las tareas que exigen una "verdad visual fuerte" continúan dependiendo de infraestructuras en la nube. La suficiencia funcional de la IA local, aunque es más soberana y respetuosa con los derechos individuales y la privacidad (Jane, 2025; Nissenbaum,

2010), resulta limitada en potencia, no está ausente de alucinaciones y supone aceptar una verdad visual limitada.

De esta manera, se puede afirmar que la suficiencia de la IA local implica aceptar una lectura del mundo más superficial. La IA local permite una gobernanza controlada por parte del usuario sobre el modelo, el relato y los datos, garantizando mayores niveles de privacidad y autonomía. En contraste, la IA visual en la nube ofrece descripciones más ricas y detalladas, cercanas a una mirada humana, pero su ejecución e interpretación dependen de infraestructuras corporativas sujetas a políticas privadas de moderación y control que escapan al ámbito de decisión del usuario. Esto introduce una forma de dependencia epistémica indirecta donde la IA en la nube redefine lo que se puede considerar conocimiento verdadero (Foucault, 1975). Todo ello, se relaciona con lo que expresan Pasquale (2015), Rouvroy y Berns (2018), sobre la producción y manipulación de verdades algorítmicas y la opacidad de las infraestructuras en la IA en la nube. Por tanto, la IA local podría ser una opción para aquellos usuarios, organizaciones y entidades que prioricen la autonomía y la privacidad, con la contrapartida de asumir la existencia de errores estructurales derivados de esa menor capacidad perceptiva. Esto es especialmente relevante para aquellas profesiones que utilizan las informaciones digitales como espacios de conocimiento de la realidad social, tal y como podría ser el periodismo.

El estudio también muestra que el control paramétrico local representa una herramienta central para la autonomía del usuario, lo cual se relaciona con la gobernanza algorítmica. La posibilidad de ajustar parámetros como top-p, top-k o la temperatura permite reducir alucinaciones y aumentar la literalidad, mejorando la trazabilidad y la previsibilidad del modelo. Como se ha expuesto en el ejemplo de la figura 4, el ajuste paramétrico reduce alucinaciones, incrementa la literalidad y hace el comportamiento del modelo más previsible y trazable. Esta capacidad se alinea con los principios de una gobernanza responsable de la información orientada a proteger autonomía y evitar daños por opacidad técnica (Floridi, 2019; Pasquale, 2015) entendiéndose como una forma de soberanía tecnológica situada (Bauwens, 2005; Mazzucato, 2021) No obstante, este control regula no solo lo que el modelo puede decir, sino cómo puede decirlo. Cabe añadir, además, que estos efectos no son accidentales, sino una consecuencia de una configuración que restringe el espacio de generación y orienta el modelo hacia salidas más previsibles y verificables, transformando al usuario en un agente activo en la modulación del sistema. Así, la nube concentra poder técnico y económico y reduce la autonomía práctica de los usuarios, reforzando asimetrías de control (Couldry y Mejías, 2019; Zuboff, 2019) y como una vía de resistencia parcial a la dependencia de infraestructuras privadas y opacas asociada al "solucionismo" (Morozov, 2013). En este sentido, se confirma la hipótesis en la que la IA local ofrece mayores niveles de transparencia y control, aunque a costa de una menor densidad interpretativa. Por ello, se distinguen dos regímenes de gobernanza: una gobernanza local, controlable y contextualizable; y una gobernanza en la nube, mayoritariamente opaca y no parametrizable por el usuario final porque gobernar los parámetros no sólo determina qué puede "ver" la IA, sino, sobre todo, cómo puede narrarlo, delimitando los márgenes de su producción discursiva.

Finalmente, se han identificado tareas que pueden resolverse eficazmente mediante IA local, lo que reduce la dependencia de la nube y permite evitar infraestructuras centralizadas de alto consumo energético. Principalmente, se trataría de: (1) el etiquetado general; (2) resúmenes de indexación; (3) títulos descriptivos; y (4) descripciones atmosféricas en escenas de baja complejidad. En estos usos, la ejecución local puede reducir la dependencia de la nube, contribuyendo a disminuir el uso de infraestructuras centralizadas de alto consumo energético y alto impacto material. Esto se alinea con el marco de ecología computacional, que subraya la materialidad de la IA y sus costes ambientales y sociales: centros de datos, cadenas extractivas, consumo energético y trabajo invisibilizado (Crawford, 2021; Ensmenger, 2021). En la medida en que parte del procesamiento pueda desplazarse a dispositivos locales o entornos controlados y a menor escala, se abre un espacio para configuraciones más sostenibles y gobernables (Moosman, 2025; Schwartz et al., 2020), especialmente si se reserva la nube para tareas de alta exigencia perceptiva.

A partir de los resultados observados, la investigación propone una lectura integradora mediante la concepción de la IA como sistema integral de relaciones, flujos y procesos (IA sistémica). Los modelos locales pueden entenderse como ecosistemas de gobernanza, sostenibilidad y privacidad, mientras que los modelos en la nube constituyen ecosistemas de precisión y densidad visual (Crawford, 2021; OECD, 2022; Strubell et al., 2019). Esta investigación no propone la sustitución de un paradigma por otro, sino una complementariedad estratégica. Los sistemas híbridos permiten preservar la autonomía y la

trazabilidad en contextos sensibles, al tiempo que recurren a la nube únicamente cuando la complejidad perceptiva lo exige. Así, la elección entre soluciones locales, en la nube o híbridas no supone una decisión meramente técnica, sino que es una decisión política, epistémica y ecológica, donde dialogan precisión, autonomía, ecología y poder interpretativo. Es por ello que se propone el concepto de EcoIA para encapsular un enfoque de diseño, evaluación y uso de los sistemas de la IA que considera de manera integrada sus dimensiones técnicas, materiales, energéticas y de valores.

No obstante, este estudio no está exento de limitaciones. El análisis se ha centrado en un conjunto acotado de modelos de IA visual, concretamente Qwen3-VL-2B como representante de IA local y Qwen3-VL-235B y Qwen3-MAX como modelos en la nube. Aunque estos modelos permiten observar con claridad las asimetrías entre arquitecturas locales y centralizadas, los resultados no pueden generalizarse automáticamente a otros modelos multimodales, especialmente aquellos con arquitecturas híbridas, entrenamiento especializado o capacidades intermedias. Futuros estudios podrían ampliar el análisis a otros modelos y arquitecturas, incluyendo sistemas multimodales de tamaño intermedio, soluciones optimizadas para *edge computing* y configuraciones híbridas que combinen procesamiento local y en la nube, con el fin de refinar la noción de suficiencia funcional y delimitar con mayor precisión los umbrales entre autonomía y dependencia centralizada. Asimismo, resulta necesario incorporar *datasets* más amplios y culturalmente diversos para examinar cómo las asimetrías perceptivas se entrelazan con sesgos contextuales y producen distintas “verdades visuales”.

La relevancia de esta investigación reside en ofrecer un análisis comparativo de la IA visual local y en la nube que supera la evaluación puramente técnica del rendimiento para situar las diferencias en un marco epistémico, político y ecológico integrado. El estudio introduce la noción de suficiencia funcional, mostrando cómo ciertas tareas visuales se neutralizan a nivel lingüístico y ocultan asimetrías perceptivas profundas, lo que resulta clave para comprender los límites reales de la autonomía de la IA local. Además, vincula el control paramétrico con formas de gobernanza algorítmica más transparentes y contextualizables, subrayando que gobernar parámetros implica gobernar el relato y no necesariamente la percepción. Al proponer el marco de la IA como sistémica y el concepto de EcoIA, la investigación aporta un enfoque original y socialmente relevante que conecta percepción visual, sostenibilidad material, autonomía del usuario y distribución del poder sociotécnico, contribuyendo así a debates actuales sobre ética, gobernanza y sostenibilidad de la inteligencia artificial.

6. Agradecimientos

El presente texto surge del proyecto Erasmus+Project: Facing Global Challenges - Environmental and Peace Education in Interreligious Cooperation for Religious Teacher Training (2023-1-DE01-KA220-HED-000166843).

7. Referencias

Bauwens, M. (2005). The Political Economy of Peer Production. *CTheory*. https://journals.uvic.ca/index.php/ctheory/article/view/14464/5306

Benjamins, R. y Salazar, I. (2020). Towards a framework for understanding societal and ethical implications of Artificial Intelligence. arXiv:2001.09750. https://doi.org/10.48550/arXiv.2001.09750

Blanco, I., Fernández-Martínez, L.M. y Suárez-Álvarez, R. (2019). *Vulnerabilidad y cultura digital: Riesgos y oportunidades de la sociedad hiperconectada*. Dykinson, S.L. https://doi.org/10.2307/j.ctv153k50r

Bratton, B. (2015). *The Stack: On Software and Sovereignty*. MIT Press.

Couldry, N., y Mejias, U. A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating it for Capitalism*. Stanford University Press.

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Bratton, B. H. (2016). *The stack: On software and sovereignty*. MIT Press.

Da Silva Barros, T., Giroire, F., Aparicio-Pardo, R. y Moulierac, J. (2025). Small is Sufficient: Reducing the World AI Energy Consumption Through Model Selection. arXiv. https://doi.org/10.48550/arXiv.2510.01889

Ensmenger, N. (2012). *The computer boys take over: Computers, programmers, and the politics of technical expertise*. MIT Press.

Eubanks, V. (2020). *Digital dead end: Fighting for social justice in the information age*. MIT Press.

European Parliamentary Research Service. (2025). *Energy and artificial intelligence: Understanding the energy implications of AI and data centres (Briefing No. 775859).* European Parliament. https://bit.ly/3NFTvUA

Federació Catalana de Fotografia (2023). *Definicions categories de fotografia de la Federació Catalana de Fotografia*. https://bit.ly/4ad9PF5

Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.

Foucault, M. (1975). *Vigilar y castigar: Nacimiento de la prisión*. Siglo XXI Editores.

Hui, Y. (2016). *The Question Concerning Technology In China*. MIT Press.

Jane, K. (2025). AI Ethics in Cloud-Based Edge Computing: Balancing Local Autonomy and Central Control. *International Journal of Advances in Engineering and Management.* https://bit.ly/4qIiLZ4

Jasanoff, S. (2016). *The ethics of invention: Technology and the human future*. W. W. Norton & Company.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. *Oxford Academic.* https://doi.org/10.1093/oso/9780199256044.001.0001

Latour, B. (1993). *We have never been modern*. Harvard University Press.

Mazzucato, M. (2021). *Mission economy: A moonshot guide to changing capitalism*. Allen Lane.

Moosmann, J., Bonazzi, P., Li, Y., Bian, S., Mayer, P., Benini, L., y Magno, M. (2025). Ultra-efficient on-device object detection on AI-integrated smart glasses with TinyissimoYOLO. Im *European Conference on Computer Vision (ECCV) 2024 Workshops* (pp. 262–280. Springer. https://doi.org/10.1007/978-3-031-91989-3_17

Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. PublicAffairs.

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

OECD (2022) Measuring the environmental impacts of artificial intelligence compute and applications: the AI footprint. OECD Publishing. https://bit.ly/4rcnMZG

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Rouvroy, A., y Berns, T. (2018). Gobernabilidad algorítmica y perspectivas de emancipación: ¿lo dispar como condición de individuación mediante la relación? Ecuador Debate, 104, 124–147. http://hdl.handle.net/10469/15424

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

Strubell, E., Ganesh, A., y McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of ACL,* 3645–3650. https://doi.org/10.18653/v1/P19-1355

Wang, X., Tang, Z., Guo, J., Meng, T., Wang, C. H., Wang, T., y Jia, W. (2025). Empowering edge intelligence: A comprehensive survey on on-device AI models. *ACM Computing Surveys, 57*(9), 1–39. https://doi.org/10.1145/3724420

Yau, S. S., y Karim, F. (2004). A context-sensitive middleware for dynamic integration of mobile devices with network infrastructures. *Journal of Parallel and Distributed Computing, 64*(2), 301–317. https://doi.org/10.1016/j.jpdc.2003.10.007

Zhang, C, Patras, P y Haddadi, H. (2019). Deep Learning in Mobile and Wireless Networking: A Survey. *IEEE Communications Surveys & Tutorials, 21*(3), 2224-2287. https://doi.org/10.1109/COMST.2019.2904897

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.