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

ARTIFICIAL INTELLIGENCE BEYOND THE CLOUD
Ethics, decentralization, Digital Ecology and Algorithmic Governance of Local Artificial Intelligences

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

KEYWORDS

AI ethics, algorithmic governance
digital power
digital sustainability
local artificial intelligence
technological decentralization

ABSTRACT

This study analyzes how visual AI models running on home devices reconfigure algorithmic governance, security, and the distribution of power. From a technological democratization perspective, it focuses on everyday computers, without high-end internal or external GPUs, to assess an AI that is energy-efficient, ethically governable, and less dependent on corporate platforms. Forgoing GPUs avoids replacing hardware and reduces environmental impact. A qualitative comparative methodology is adopted, based on controlled experiments with local and cloud-based visual AI models applied to various image types, with systematic variations in algorithmic parameters. The results show that local visual AI, although limited in performance, is sufficient for descriptive tasks. Its use, governed through algorithmic control, can redistribute technological power and offer alternatives to cloud centralization, strengthening autonomy, privacy, and institutional capacity.

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

1. Introduction

Although artificial intelligence (AI) systems have been deployed very quickly, their development has focused mainly on large-scale models running in the cloud, prioritising performance and scalability metrics over control, sustainability, content circulation or governance (OECD, 2022; Strubell et al., 2019). This process has contributed to increased computational centralisation, with high energy costs and a growing concentration of power in the hands of large technology platforms that control access, control and content circulation (Crawford, 2021). Authors such as Schwarts et al. (2019) have also questioned the environmental impact of these large-scale models and their dependence on energy-intensive infrastructure.

AI systems running locally should be understood as generative AI models downloaded and executed from domestic devices, which will be referred to as local AI from now on. Compared to cloud models, these models not only guarantee data privacy, but also redistribute algorithmic power, enable situated ethical governance, reinforce institutional autonomy, and reduce both corporate dependence and ecological impact. Wang et al. (2025) identify a number of advantages of these models over those run in the cloud. Firstly, they are more sustainable, as they reduce the use of data centres and the massive and redundant transfer of information. Secondly, by remaining under the direct control of the user, data privacy and sovereignty are ensured. For example, in the case of local AI for image analysis, the images do not leave the devices, so there is no transfer to third parties or storage on corporate servers. This facilitates regulatory compliance in terms of data protection—including the General Data Protection Regulation (GDPR)—protection of minors, and sensitive environments. Additionally, there is a redistribution of technological power, as control is no longer in the hands of centralised platforms but rather in local decisions about which model to use, how to configure it, and when to apply it. This capability enables effective algorithmic governance, based on the adjustment of parameters such as temperature or top-p, where ethics ceases to be a corporate policy and becomes a local technical decision, including the application of localised ethical prompts. Finally, the use of local visual AI strengthens institutional autonomy, as it does not require licences, allows offline use and is not subject to changes in platform policies.

In this context, local AI models have been relatively unexplored and rarely considered as a technical, political, and ecological alternative to computational centralisation in the cloud, largely due to the acceleration of technological change, which has hindered a systematic and situated analysis of their potentialities and limitations. From a socio-technical perspective, this invisibility is consistent with what Latour (1993) describes as the effect of dominant infrastructures: the more they function and expand, the more opaque they become and the less visible their material and political mediations are. According to Latour, modernity is sustained by two sets of simultaneous practices: purification, which produces two distinct ontological zones: that of non-humans (nature) and that of humans (society/culture); and mediation or translation, which generates hybrids that mix the social and the material (Latour, 1993). However, the so-called modern constitution systematically conceals the work of mediation, while making only the work of purification visible and legitimate (Latour, 1993). In turn, in line with Yuk Hui (2016), technological acceleration tends to consolidate hegemonic technical cosmologies, hindering the articulation of local alternatives where regimes of energy, meaning and governance can be defined in proximity.

In this work, AI is conceptualised as a comprehensive system of relationships, flows, and impacts that emerge when the production, processing, and governance of images are reorganised outside centralised infrastructures, simultaneously addressing technical, political, energy, and ethical dimensions. It is proposed that local AI applied to image processing, while offering limited performance but sufficient for certain descriptive tasks, can redistribute technological power, reduce ecological impact and offer ethical and secure alternatives to cloud centralisation, reinforcing autonomy, privacy and governance through local algorithmic control.

Based on this framework, the research focuses on a comparative analysis of the use of visual AI in the cloud and locally applied to the creation and description of images, with the aim of examining its implications in terms of power, algorithmic governance, sustainability, and technological control. Visual AI refers to generative AI models that have the multimodal capacity to analyse and interpret images. Therefore, this paper discusses how the descriptive and interpretative behaviour of locally executed visual AI models differs from cloud-based models. This comparison is based on accuracy, stability,

hallucinations and biases, and on the types of tasks and contexts of use in which local AI is sufficient without requiring cloud infrastructure, especially in sensitive contexts. Sensitive contexts would be those in which minimal interaction or action can produce significant changes in terms of outcome (Yau and Karim, 2004). These changes could be irreversible or costly in terms of outcome.

For this reason, the main objective of this research is to analyse the behaviour of AI models applied to images executed locally and in the cloud in image description and interpretation tasks. The aim is to identify in which types of tasks and contexts of use local AI is functionally sufficient and in which cases cloud infrastructure is necessary.

As a novelty, this research comprehensively evaluates how local implementation affects the descriptive and interpretative behaviour of models, the transparency and traceability of parametric control, functional adequacy for everyday uses, the reduction of the energy footprint, and the reconfiguration of technological dependency relationships, an approach that has only been addressed in an incipient manner to date (Wang et al., 2025; Zhang et al., 2022). This transversal approach allows for a critical examination of how local visual AI can contribute to more democratic algorithmic governance, greater sustainability, and effective user autonomy.

2. Theoretical framework

Based on its architecture, control and social impact, a distinction is made between cloud AI and local AI. The former centralises data processing and storage in external infrastructures managed by large platforms, which enhances large-scale learning and control of large volumes of information, thereby limiting user autonomy. The latter operates directly on user devices or in community-controlled environments, allowing for greater data privacy. Its learning and prediction potential may be lower than in the cloud, but users remain autonomous and dependence on centralised platforms is reduced, respecting individual rights (Jane, 2025).

Visual AI has been developed primarily in the cloud, where large-scale models are trained and executed on centralised, high-performance infrastructures. This implies a heavy reliance on energy-intensive data centres, very limited operational transparency, and a concentration of technical and economic control in large platforms (Crawford, 2021; OECD, 2022; Strubell et al., 2019). Local visual AI models that are run on domestic devices are proposed as an alternative, processing data in a way that improves privacy while reducing the energy footprint and enabling more governable forms of content production and circulation because they operate with hardware constraints and on a smaller scale (Moosman, 2025; Schwartz et al, 2020).

2.1. Governance, control, and power

AI has been analysed by various authors as a device that redefines the modes of exercising power, governance and ethics in contemporary society. For example, Zuboff (2019) considers AI to be an instrument of surveillance capitalism, based on the extraction and predictive valuation of large volumes of data, which can generate power asymmetries. Along the same lines, Couldry and Mejías (2019) describe this process as data colonialism and highlight how the concentration of infrastructure and platforms reinforces existing structural inequalities. Rouvroy and Berns (2018) argue that algorithmic governmentality, through automated data processing, deactivates the political subject and warn of the replacement of collective decision-making by technical optimisation. Meanwhile, Pascuale (2015) notes the opacity of algorithms and their ability to concentrate power and control in key companies and sectors (e.g., the economy, information, or digital platforms), which reinforces the need for public transparency mechanisms to prevent abuses of power. It is for all these reasons that Bratton (2016) proposes understanding AI not only as a technology but as a critical element of power structures.

The distinction between local AI and cloud AI is not only technical, but also political, as it involves different regimes of governance, control and distribution of power. From this perspective, and in dialogue with Foucault (1975), AI would not only be a technical tool, but can be understood as a technology of power that organises, classifies and produces knowledge about individuals and societies, in line with the concept of the panopticon and the logics of disciplinary power. From this perspective, AI is not limited to repressing or excluding, but has the capacity to generate new realities, objects and truths. Thus, AI algorithms can influence the perception of reality through the selection of information,

the creation of content and automated decision-making. This implies a reconfiguration of power relations, as AI can operate autonomously and perpetuate dynamics of control without the need for direct human intervention. This raises fundamental ethical questions, as the implementation and use of AI can reinforce inequalities, perpetuate biases and affect individual autonomy, generating social vulnerabilities derived from the use of artificial intelligence (Benjamins and Salazar, 2020; Blanco et al., 2019).

Decentralisation and technological sovereignty challenge the artificial intelligence model based on large centralised platforms and cloud services. For Morozov (2013), technological solutionism reinforces dependence on private and often opaque infrastructures, limiting the ability of governments and citizens to decide on the use of data. This translates into a concentration of computational, economic and political power in very few companies and hinders the development of local or public alternatives. Mazzucato (2021) proposes more distributed forms of organisation geared towards the public interest, and Bauwens (2005) proposes the creation of digital communities and distributed governance to reduce dependence on the cloud.

2.2. Responsible technology and situated ethics

The uses of AI, both in the cloud and locally, are understood as a socio-technical practice in which its effects depend on the context in which it is designed, implemented and governed. Floridi (2019) introduces the principles for responsible information governance, which involves introducing ethical criteria into the regulation and implementation of technologies to protect people's autonomy. Along the same lines, Eubanks (2020) warns about the social inequalities that technologies such as AI reinforce, limiting access to rights and resources, and therefore also concludes that there is a need for regulation with a focus on social justice. All of this coincides with what Nissenbaum (2010) points out in his idea of contextual privacy, considering that the appropriate use of information depends on the social and regulatory context in which it is produced, and thus the design of respectful technologies should take into account privacy as compliance with social norms that regulate the circulation of information. Jasanoff (2016) points out that technological innovation is not neutral and that its impact depends on how it is designed, regulated and used in specific social contexts, and therefore proposes shared responsibility among all actors involved in technological governance.

2.3. Computational ecology and sustainability

Artificial intelligence systems are material infrastructures with a significant social and environmental impact. Crawford (2021) highlights how their operation depends on mineral extraction, high energy consumption and the generation of large volumes of electronic waste. Ensmenger (2021) adds to this the social costs of their use, such as the labour impact of maintaining data centres (often with precarious and invisible work) and mass data processing. Therefore, the use of AI also becomes an ecological issue and focuses the debate on sustainability due to the impact of its ecological and social footprint. All of this makes sense in a global context in which contemporary computing is understood, as Bratton (2015) argues, as a planetary architecture composed of multiple interrelated layers: physical infrastructure (energy, data centres, cables, satellites), platforms, users and forms of governance. Computing relies on global energy, extraction and logistics systems, so its ecological consequences are structural. In fact, data centres consumed 415 TWh of electricity in 2024, meaning that a large data centre can consume the electricity of 100,000 households in a year. And although it only accounts for 1.5% of global electricity in 2024, energy demand is growing rapidly and is expected to double by 2030 to 945 TWh (European Parliamentary Research Service, 2025). In contrast, if smaller models (such as local ones) are used, global energy consumption could be reduced by 27.8%, or 31.9% by 2025 (da Silva Barros et al., 2025).

3. Methodological framework

This research adopts a comparative qualitative methodology based on controlled experimentation, structured through peer comparisons between visual AI models running in local domestic environments and equivalent cloud-based models. To ensure the comparability of the results, the same tasks, prompts, and image types were applied in both environments, varying only the execution conditions, computational infrastructure, and degree of parametric control. This approach allows for the analysis

of model behaviour under functionally different conditions, isolating the impact of centralisation, algorithmic control, and platform dependency on the results.

An analysis was carried out on a homogeneous set of images, organised into six categories: urban scenes, natural environments, night-time images, inanimate nature (still lifes), abstract compositions (light painting) and dynamic sports scenes, using the definitions of categories provided by the Catalan Photography Federation (2023) as a reference. Each image has been processed by all models following an identical prompt protocol, consisting of three sequential tasks:

- Prompt for factual visual description: 'Describe with extreme precision all visible objects in the image, including colours, shapes, and spatial relationships. Do not invent anything'.
- Prompt for interpreting the scene and action: 'Explain what is happening in the image, what action the people or objects are performing, and what elements allow you to deduce it'.
- Prompt for brief summarization and title generation: 'Summarize the main content of the image in three clear sentences. Then generate three alternative titles without adding unnecessary details'.

For the analysis of the local environment, small and medium-sized multimodal visual AI models were used, running on mid-range domestic devices via local platforms, specifically: Qwen2.5-VL-7B-Instruct, with standard parameters (Top P 0.9, Top k 40, Temperature 0.6) and Qwen3-VL-2B-Instruct with restrictive parameters (Top-p 0.5; Top-k 15; Temperature 0.05). These models were selected for their viability on domestic hardware without high-end GPUs. In the cloud environment, larger-scale models with greater visual capacity were used, specifically: Qwen3-VL-32B and Qwen3-VL-235B-A22B. Systematic variations in algorithmic parameters (temperature, top-p, top-k) were introduced into the local models in order to observe their impact on stability, hallucinations, biases and interpretative control. The cloud models were analysed with their default settings, reflecting the actual conditions of use on centralised platforms.

Outcomes were evaluated comparatively based on parameters of descriptive accuracy; output stability (e.g., the occurrence of hallucinations, semantic control, and process traceability); and suitability for variable and sensitive contexts (different images with different characteristics, such as the representation of dynamic scenes). Criteria of functional convenience (it can do what the user asks it to do), energy efficiency (it does not depend on external resources such as the cloud) and governance (the user can control it) have also been applied. Descriptive accuracy is understood as the model's ability to describe what actually appears in the image (e.g. correct identification of visible elements, characteristics and relationships) without adding information that is not present. Algorithmic governance is defined here as the ability to control the model locally, even without a network connection, ensuring autonomy, technological sovereignty and security by not outsourcing data to the cloud.

The first and second authors used the same technical specifications for the analysis, including computer equipment, AI models (local and cloud-based), prompts, and images. Once the AI analysis was returned, the results obtained in the experiments used were systematically compared and contrasted. This design makes it possible to identify the types of tasks for which local AI is sufficient and when the use of cloud models provides significant advantages, providing evidence for informed decision-making about local, cloud or hybrid configurations.

4. Analysis

The research presents a comparative analysis of the computer vision capabilities of local AI models (Qwen3-VL-2B) and cloud-based AI models (Qwen3-VL-235B and Qwen3-MAX), using a diverse dataset of six image types: (1) urban architecture (Figure 1); (2) nature (Figure 2); (3) night scenes (Figure 3); (4) still lifes (Figure 4); (5) abstract art (Figure 5); and sports action (Figure 6), according to the classification of the Catalan Photography Federation (2023).

Figure 1. Photograph used for urban architecture typology.



Source: iStock/nito103, 2011

Figure 2. Photograph used for nature typology.



Source: iStock/FG Trade, 2018

Figure 3. Photography used for night scenes typology.



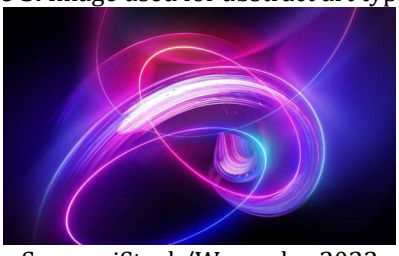
Source: iStock/Florian Parzefall, 2020

Figure 4. Photograph used for still lifes typology.



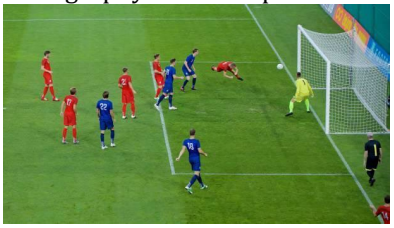
Source: iStock/RCerruti, 2012

Figure 5. Image used for abstract art typology.



Source: iStock/Wacomka, 2022

Figure 6. Photography used for sports action typology.



Fuente: iStock/Simonkr, 2017

The experiments yielded a large amount of data. For reasons of length and clarity, only those results that are noteworthy from the point of view of effectiveness and efficiency in relation to description, brief synthesis, and image caption generation will be presented. The results are therefore structured around three axes: (1) visual descriptive accuracy; (2) performance in information synthesis; (3) performance according to the semantic and dynamic complexity of the images.

4.1. Visual descriptive accuracy

A hierarchy can be observed in the algorithmic 'gaze'. Local models operate primarily on the basis of a general and superficial perception, while cloud models achieve a structural or expert perception. Local models correctly identify basic categories, but fail to capture the volumetric structure or technical details of the scene presented. Therefore, they tend to resort to replacing hallucinatory features, filling in perceptual gaps with internal prototypes. This results in critical errors, such as the incorrect identification of the Arc de Triomphe (Figure 1) as the 'Arc de la Mercè' or the invention of non-existent architectural elements. In contrast, cloud models demonstrate far superior structural visual competence, perceiving materials (aged wood, rusted metal), precise architectural geometries, and complex natural processes, such as river erosion. In images of high visual complexity, for example, light painting (Figure 5), they are the only ones capable of correctly reconstructing the three-dimensional geometry of light trails.

When restrictive parameters are applied (top-p 0.5, top-k 15, T=0.05), they act as a containment mechanism: they reduce hallucinations and improve literalness, but do not increase the model's actual perceptual capacity. With these parameters, the local model stops inventing, but still fails to 'see' fine detail. For example, in Figure 4, the local model goes from describing 'The architectural elements (brick vaulted arches, stone or brick supports for the barrels, and the general layout) suggest a historic or traditional-style cellar [...]' to a more literal output such as 'The image shows an underground cellar with a long corridor lined with large wooden barrels. The space is constructed with arched brick walls and a concrete floor." In contrast, the MAX model performs an in-depth architectural reading of the cellar, detecting the texture of aged wood, rust on metals, and the rhythmic symmetry of the brick arches. In the example of the Arc de Triomf, the local model sees a monument, but often invents side arches or gets the name wrong. In contrast, centralised models accurately perceive the friezes, sculptures, and depth of the urban ensemble.

Thus, the use of AI in the cloud means that image interpretation depends on corporate servers that apply moderation filters and censorship policies over which the user has no control. However, a local model allows the user to adjust and control output parameters such as temperature or top-p to obtain a strictly literal description, guaranteeing the privacy and autonomy of the narrative, even if the visual capacity is lower.

4.2. Performance in summary tasks

A noteworthy finding is that the summarisation task is not discriminatory in assessing the visual analysis capabilities of the models, as the same result is obtained. In the six images analysed, the differences in the extreme description disappear in the synthesis phase. All models manage to produce functionally equivalent summaries, which equalise or mask the vision deficiencies of the smaller models.

In tasks involving the identification of basic categories and entities, local models perform comparably to cloud models. In static images with low semantic complexity, such as still lifes or architecture, both correctly identify animate and inanimate static bodies, such as people, animals and objects, as well as the general nature of the environment (natural, urban or interior spaces). However, in night-time images analysed locally, which involve low lighting, ambiguity increases, detail is lost and there are more hallucinations.

Secondly, summarisation and indexing tasks, which constitute what can be called a 'neutralisation zone', do not allow for differentiation between the visual quality of the models. When the goal is to generate a brief informative description (e.g., a three-sentence summary or a descriptive title), the local model is as useful as the cloud model. Although the latter often provides more refined stylistic or terminological nuances, the basic communicative function is fulfilled in both cases.

Likewise, in static nature scenes (Figure 2) or night-time images without complex action (Figure 3), the performance gap is limited. The local model correctly identifies fundamental elements such as rivers, vegetation, sky, silhouettes, or reflections, allowing for the construction of basic atmospheric descriptions. Although it is not capable of recognising geomorphological processes or advanced lighting dynamics, its performance is sufficient for general descriptive uses. In this context, local models also show an adequate ability to describe colour palettes and global lighting, facilitating an elementary aesthetic classification of the image.

Conversly, cloud models have significantly superior structural visual capabilities, allowing them to capture technical and architectural details with high precision. This density of information can lead to over-analytical interpretations, emphasising textures, symmetries or physical processes that are not always necessary for a simple description. While local models are limited to a general perception, dense models analyse spatial structure, the physics of light or materials in depth, which can generate information overload for users who are only looking for a basic description.

These results show that, although cloud models 'perceive' much more than local models, this advantage is neutralised in short synthesis tasks, where visual granularity is not a requirement. Consequently, the summaries generated by both types of models tend to be functionally equivalent.

The adequacy of the local model is surpassed when the task requires nominal precision, technical or tactical interpretation (as in sports scenes), or volumetric and spatial understanding. In these cases, local models lose reliability, while cloud models maintain robust visual interpretation.

Thus, local models are suitable for general labelling, summarisation, indexing, and contexts where governance, privacy, and user autonomy are priorities. However, for tasks requiring strong visual veracity or expert analysis, reliance on cloud models remains necessary. The sources analysed agree that the perceptual superiority of the latter is evident in structured perception tasks, although it is often hidden or unnecessary in superficial textual tasks.

4.3. Semantic complexity and image dynamics

The results show that the magnitude of the difference between local models and cloud models is not uniform, but varies significantly depending on the semantic and dynamic complexity of the visual content analysed. The architecture of the model directly affects its structural visual competence, generating a performance gap that widens depending on the nature of the scene.

For example, in static or natural scenes, such as daytime landscapes, winding rivers, or night-time images without complex action, the difference between the two types of models is moderate. In these contexts, local models are sufficient for basic identification of the elements present in the image. Their performance is based on the identification of global categories (such as water, vegetation, sky, or human silhouettes), which allows for the construction of functional descriptions, albeit lacking in structural depth.

This is explained by the use of prototype-based identification mechanisms, also described as hallucinated feature substitution. In scenes with low semantic complexity, the local model recognises general patterns and completes the description using standardised internal representations. Since the correction criterion is eminently categorical (e.g., identifying whether it is a river or a night scene), perceptual errors are less obvious to the user. Furthermore, in this type of image there is no 'rule' or 'tactic' to interpret: the task is fundamentally linguistic and atmospheric, allowing the local model to describe colour gradients, general lighting or the arrangement of elements without the need to understand geomorphological processes or deep photographic structures, capabilities that denser models do exhibit. Conversely, in dynamic and socially structured scenes, such as sporting events, the gap between local and cloud models reaches its maximum expression. The case of football (Figure 6) is a paradigmatic example of a structured perceptual task, as the correct interpretation of the image requires a spatial, kinematic and tactical reconstruction that goes far beyond the simple identification of objects.

In this context, it can be observed that cloud models (235B and MAX) act in a similar way to a professional VAR analyst. They accurately analyse the body orientation of players, escape lines, the trajectory of the ball and the spatial relationship between the different elements of the scene. This ability allows them to determine whether the ball has crossed the goal line or whether the goalkeeper has arrived late to make the save, inferences that require high-resolution multimodal representation. In contrast, the local model is limited to flat, generic descriptions, such as "a player kicks the ball", without

identifying specific technical actions or interpreting the tactical situation. This limitation leads to critical misinterpretations, such as stating that the goalkeeper has stopped the ball when the image shows that he has been beaten, or introducing non-existent static plays (such as a “free kick”) in the middle of a dynamic action.

5. Discussion and conclusions

The aim of this study was to identify the types of tasks and contexts of use in which local AI is functionally sufficient and those in which cloud infrastructure is necessary. Results highlight that observable differences between locally executed visual AI models and cloud-based models cannot be interpreted solely in terms of technical performance, but must be understood as epistemic asymmetries: what kind of ‘visual truth’ (i.e., what is assumed by most to be evident, accurate, and verifiable) is produced and with what reliability; and algorithmic governance (who controls the infrastructure, parameters, data and, ultimately, the narrative). Furthermore, these differences have political, ecological and social implications (Floridi, 2019; Jasanoff, 2016).

The first finding observed is that local models have conditional functional sufficiency. On the one hand, they are capable of performing basic descriptive tasks, such as general labelling, scene classification, or the generation of brief summaries. For example, nominal and structural hallucinations are observed, such as confusing Arc del Triomf with Arc de la Mercè. In other words, they are unable to identify volumetric structures and lack nominal or spatial dynamic precision, making them vulnerable to errors and the phenomenon of hallucinatory feature substitution. It can therefore be deduced that local AI is adequate or sufficient for contexts of low semantic and dynamic complexity, while tasks that require ‘strong visual truth’ continue to depend on cloud infrastructures. The functional sufficiency of local AI, although more sovereign and respectful of individual rights and privacy (Jane, 2025; Nissenbaum, 2010), is limited in potential, is not free from hallucinations, and implies accepting a limited visual truth.

Thus, it could be said that the adequacy of local AI implies accepting a more superficial reading of the world. Local AI allows for user-controlled governance over the model, narrative, and data, ensuring greater levels of privacy and autonomy. In contrast, cloud-based visual AI offers richer and more detailed descriptions, closer to a human perspective, but its execution and interpretation depend on corporate infrastructures subject to private moderation and control policies that are beyond the user's control. This introduces a form of indirect epistemic dependence where cloud AI redefines what can be considered true knowledge (Foucault, 1975). All of this is related to what Pasquale (2015), Rouvroy and Berns (2018) express about the production and manipulation of algorithmic truths and the opacity of infrastructures in cloud AI. Therefore, local AI could be an option for those users, organisations and entities that prioritise autonomy and privacy, with the trade-off of accepting the existence of structural errors derived from this reduced perceptual capacity. This is especially relevant for professions that use digital information as a means of understanding social reality, such as journalism.

The study also shows that local parametric control is a key tool for user autonomy, which is related to algorithmic governance. The ability to adjust parameters such as top-p, top-k, or temperature reduces hallucinations and increases literality, improving the traceability and predictability of the model. As shown in the example in Figure 4, parametric adjustment reduces hallucinations, increases literality, and makes the model's behaviour more predictable and traceable. This capability aligns with the principles of responsible information governance aimed at protecting autonomy and avoiding harm due to technical opacity (Floridi, 2019; Pasquale, 2015), understood as a form of situated technological sovereignty (Bauwens, 2005; Mazzucato, 2021). However, this control regulates not only what the model can say, but how it can say it. It should also be added that these effects are not accidental, but rather a consequence of a configuration that restricts the space for generation and directs the model towards more predictable and verifiable outputs, transforming the user into an active agent in the modulation of the system. Thus, the cloud concentrates technical and economic power and reduces the practical autonomy of users, reinforcing asymmetries of control (Couldry and Mejías, 2019; Zuboff, 2019) and as a means of partial resistance to the dependence on private and opaque infrastructures associated with ‘solutionism’ (Morozov, 2013). In this sense, the hypothesis that local AI offers greater levels of transparency and control is confirmed, albeit at the cost of lower interpretative density. Therefore, two governance regimes can be distinguished: local governance, which is controllable and

contextualisable; and cloud governance, which is mostly opaque and cannot be parameterised by the end user because governing the parameters not only determines what the AI can 'see' but, above all, how it can narrate it, delimiting the margins of its discursive production.

Lastly, tasks have been identified that can be effectively solved using local AI, reducing dependence on the cloud and avoiding energy-intensive centralised infrastructures. These mainly include: (1) general labelling; (2) indexing summaries; (3) descriptive titles; and (4) atmospheric descriptions in low-complexity scenes. In these uses, local execution can reduce dependence on the cloud, helping to decrease the use of centralised infrastructures with high energy consumption and high material impact. This aligns with the computational ecology framework, which highlights the materiality of AI and its environmental and social costs: data centres, extractive chains, energy consumption, and invisible labour (Crawford, 2021; Ensmenger, 2021). To the extent that part of the processing can be shifted to local devices or controlled, smaller-scale environments, this opens up space for more sustainable and governable configurations (Moosman, 2025; Schwartz et al., 2020), especially if the cloud is reserved for tasks with high perceptual demands.

Based on the results observed, the research proposes an integrative interpretation by conceiving AI as a comprehensive system of relationships, flows, and processes (systemic AI). Local models can be understood as ecosystems of governance, sustainability and privacy, while cloud models constitute ecosystems of precision and visual density (Crawford, 2021; OECD, 2022; Strubell et al., 2019). This research does not propose the replacement of one paradigm with another, but rather a strategic complementarity. Hybrid systems make it possible to preserve autonomy and traceability in sensitive contexts, while resorting to the cloud only when perceptual complexity requires it. Thus, the choice between local, cloud or hybrid solutions is not merely a technical decision, but a political, epistemic and ecological one, where precision, autonomy, ecology and interpretative power come into play. This is why the concept of EcoIA is proposed to encapsulate an approach to the design, evaluation and use of AI systems that considers their technical, material, energy and value dimensions in an integrated manner.

However, this study is not without limitations. The analysis focused on a limited set of visual AI models, specifically Qwen3-VL-2B as representative of local AI and Qwen3-VL-235B and Qwen3-MAX as cloud models. Although these models clearly reveal the asymmetries between local and centralised architectures, the results cannot be automatically generalised to other multimodal models, especially those with hybrid architectures, specialised training or intermediate capabilities. Future studies could extend the analysis to other models and architectures, including medium-sized multimodal systems, solutions optimised for edge computing, and hybrid configurations that combine local and cloud processing, in order to refine the notion of functional sufficiency and more precisely delineate the thresholds between autonomy and centralised dependency. It is also necessary to incorporate broader and more culturally diverse datasets to examine how perceptual asymmetries intertwine with contextual biases and produce different 'visual truths.'

The relevance of this research stems from its comparative analysis of local and cloud-based visual AI, which goes beyond a purely technical assessment of performance to place the differences within an integrated epistemic, political and ecological framework. The study introduces the notion of functional sufficiency, showing how certain visual tasks are neutralised at the linguistic level and conceal deep perceptual asymmetries, which is key to understanding the real limits of local AI autonomy. Furthermore, it links parametric control to more transparent and contextualisable forms of algorithmic governance, emphasising that governing parameters implies governing the narrative and not necessarily perception. By proposing the framework of AI as systemic and the concept of EcoAI, the research provides an original and socially relevant approach that connects visual perception, material sustainability, user autonomy and the distribution of socio-technical power, thus contributing to current debates on the ethics, governance and sustainability of artificial intelligence.

6. Acknowledgements

The present work emerges from the project Erasmus+Project: Facing Global Challenges - Environmental and Peace Education in Interreligious Cooperation for Religious Teacher Training (2023-1-DE01-KA220-HED-000166843).

7. Bibliography

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

Benjamins, R. & 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. & Suárez-Álvarez, R. (2019). *Vulnerabilidad y cultura digital: Riesgos & 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., & 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. & 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., & 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., & 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., & 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. *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., & 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., & 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 & 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.