Generative Artificial Intelligence integrated into the digital ecosystem

A framework for algorithmic governmentality

Inteligencia Artificial Generativa integrada al ecosistema digital

Un marco de situación para la gubernamentalidad algorítmica

Inteligência Artificial Generativa integrada ao ecossistema digital

Uma estrutura de situação para a governamentalidade algorítmica

DOI: https://doi.org/10.18861/ic.2025.20.1.3931

▶ AGUSTINA LASSI

alassi@unlam.edu.ar– Buenos Aires – Universidad Nacional de La Matanza, Argentina.

ORCID: https://orcid.org/0000-0003-3171-6258

HOW TO CITE: Lassi, A. (2025). Generative Artificial Intelligence integrated into the digital ecosystem. A framework for algorithmic governmentality. *InMediaciones de la Comunicación, 20*(1). DOI: https://doi.org/10.18861/ic.2025.20.1.3931

Submission date: December 20, 2024
Acceptance date: February 19, 2025

ABSTRACT

This article analyzes the advancement of the integration of Generative Artificial Intelligences into platforms, especially those based on Large Language Models (LLMs). These changes are beginning to alter the way content is produced and used, information is searched and processed, and user profiles are managed. The aim of the article is to map the main LLM developments in the West and their strategic, hardware supply

or financial partnerships with cloud computing platforms, hardware producers and servers. In this sense, we start with the widespread release of ChatGPT as the founding moment of this new stage and go through the main uses and applications of transformers in platform environments. Although we are at the beginning of the application of this technology in an integrated manner, it could generate a deepening of algorithmic personalization, modifying the forms of human subjectivation and favoring epistemic bubbles at the cognitive level, as well as concentrating supply and development at the level of political economy. That is why it is necessary to seek more forms of human intervention in data curation and to increase the active observance –both civil and governmental– of these systems and their alliances in order to avoid extreme concentration and to attend to the biases that could generate negative effects on culture.

KEYWORDS: *artificial intelligence, LLMs, algorithmic governmentality, platforms, epistemic bubbles.*

RESUMEN

Este artículo analiza el avance de la integración de las Inteligencias Artificiales Generativas a las plataformas, sobre todo la que se basan en los modelos de lenguaje grandes (LLMs). Estos cambios comienzan a alterar la manera en que se produce y utiliza contenido, se busca y procesa información, y se gestionan los perfiles de los usuarios. El objetivo del artículo es realizar un mapeo de los principales desarrollos de LLMs en Occidente y sus asociaciones estratégicas, de hardware supply o financieras con plataformas, productores de hardware y servidores de cloud computing. En tal sentido, se parte del lanzamiento masivo de Chat GPT como momento fundacional de esta nueva etapa y se recorren los principales usos y aplicaciones de transformers en los entornos platafórmicos. Si bien estamos atravesando los comienzos de la aplicación de esta tecnología de manera integrada, podría generar una profundización en la personalización algorítmica, modificando las formas de subjetivación humanas y favoreciendo las burbujas epistémicas en el plano cognoscitivo, además de concentrar la oferta y el desarrollo en el plano de la economía política. Es por eso, que se plantea la necesidad de buscar más formas de intervención humana en la curación de datos e incrementar la observancia activa –tanto civil como gubernamental– de estos sistemas y sus alianzas para evitar la concentración extrema y atender a los sesgos que pudieran generar efectos negativos en la cultura.

PALABRAS CLAVE: *inteligencia artificial, LLMs, gubernamentalidad algorítmica, plataformas, burbujas epistémicas.*

RESUMO

Este artigo analisa o avanço da integração das Inteligências Artificiais Generativas nas plataformas, nomeadamente nas que se baseiam em Modelos de Linguagem de Grande Dimensão (MLG). Estas mudanças começam a alterar a forma como os conteúdos são produzidos e utilizados, a informação é pesquisada e processada e os perfis dos utilizadores são geridos. O objetivo do artigo é mapear os principais desenvolvimentos de LLM no Ocidente e as suas parcerias estratégicas, de fornecimento de hardware ou financeiras com plataformas de computação em nuvem, produtores de hardware e servidores. Neste sentido, começa com o lançamento massivo do ChatGPT como momento fundador desta nova etapa e passa pelas principais utilizações e aplicações dos transformadores em ambientes de plataforma. Embora estejamos no início da aplicação desta tecnologia de forma integrada, ela poderá gerar um aprofundamento da personalização algorítmica, modificando as formas de subjetivação humana e favorecendo bolhas epistémicas ao nível cognitivo, bem como concentrando a oferta e o desenvolvimento ao nível da economia política. Isto suscita a necessidade de procurar mais formas de intervenção humana na curadoria de dados e de aumentar o controlo ativo, civil e governamental, destes sistemas e das suas alianças, a fim de evitar uma concentração extrema e de abordar os enviesamentos que podem gerar efeitos negativos na cultura.

PALAVRAS-CHAVE: *inteligência artificial, LLMs, governação algorítmica, plataformas, bolhas epistémicas.*

1. INTRODUCTION

The development of artificial intelligence (AI) algorithms for language understanding and generation has been a key focus of major technology corporations in both the East and West over the past two decades. This progress has been made possible by massive access to training data and the increasing processing power provided by graphics processing units (GPUs)[1]. Thus, there has been significant progress in research into language modeling, moving from *statistical models* to *neural models*. Recently, pre-trained language models have been introduced (PLM, for its acronym in English) using *transformers* on large corpora, showing a high capacity for various natural language processing tasks (NLP, for its acronym in English). Scaling up models has been proven to improve their performance. By increasing the number of parameters, they not only significantly improve in performance, but also acquire special abilities (for example, contextual learning), something that does not happen in smaller models. It is this contextual learning capability that generates a sort of "shift in the tectonic plates" regarding what the new capabilities of platforms that integrate AI into their operation might mean.

Since the massive release of ChatGPT-3.5 in November 2022 by *Open AI*, *Large Language Models* (LLMs) have been the subject of fascination and experimentation beyond the academic and scientific realm. The launch of *Chat GPT* lowered the barriers to entry for conversational *chatbots* and other forms of AI, marking the beginning of a new era in which the term "artificial intelligence" became commonplace. Behind this technology are *transformers*, the architecture that made its development possible. A *transformer* model is a neural network that learns context and meaning by tracking relationships in sequential data. It applies a set of mathematical, self-aware techniques to detect subtle ways in which data elements in a series influence and depend on each other. LLMs, or large language models, are the key component behind text generation. They consist of transformers pre-trained to predict, given an input text, the next word (or, more precisely, *token*). Since language models predict one *token* at a time, generating complete sentences requires a more elaborate approach than simply calling them once. To do this, autoregressive generation is used, a procedure in which the model is called iteratively, using its own previous results as input to continue the sequence and improve the coherence of the text. In these cases, the model generates data sequences – such as text – using its own previous outputs as new inputs, starting with an initial data set. This process is repeated iteratively to produce a complete sequence. The machine learning potential associated with these models is endless and will generate changes in the way we understand our environment and generate knowledge.

1 The acronym GPU – or graphics processing unit – refers to processors designed to handle and accelerate graphics and image processing in devices such as video cards, motherboards, mobile phones, and personal computers. By performing mathematical calculations quickly, it reduces the amount of time it takes for a computer to run multiple programs, making it an essential enabler of emerging and future technologies such as machine learning (ML), AI, and blockchain.

Transformers were first introduced in 2017 at the *NeurIPS* conference through *Google's* white paper titled *"Attention is all you need* ", kicking off a wave of machine learning advancements that some are calling "*Transformer AI*"[2]. *Generative Pre-trained Transformers* (GPTs) are natural language models that use deep neural networks to process and generate text. They are designed to understand and produce human language in a similar way to a human being, although their knowledge comes from large data sets. The most widely used GPT, *ChatGPT*, has over 200 million weekly active users – as of November 2024 – and has a simple interface in which the user (in its "free" version) can access the history of chat interactions from the last 7 days, view sample conversations, or enter text with the *prompt* (instruction or text used to interact with the AI), either by voice or by keyboard. Some of these organizations and/or companies are dedicated exclusively to strengthening and further training these language models, while others use them as an *Application Programming Interface* (API) to develop new products and services. In the West, there are currently more than 7,000 companies and organizations developing *software* applications, and *hardware* with integrated AI. The areas in which these companies operate are varied, including: logistics, finance, insurance, credit management, search, resale, education, aerospace and defense, healthcare, customer service, cybersecurity, *gaming*, waste management, *information technology* (IT)[3] and DevOps[4], automation, agriculture.

This article will examine three levels of dynamics in the advancement of these technologies. The first seeks to document the integration of Generative Artificial Intelligences of language and images into *software and hardware*, as well as platforms, and within this framework will reflect on the concept of the *publishing agency*. On a second level, we will analyze the most recognized and thriving LLM *startups* in the United States and other countries, as well as their connections with *hardware* manufacturing platform companies and *cloud computing* services, to observe the links between these *startups* and the corporations that concentrate investments in these developments. Finally, the third level aims to delve into the types of links (interoperability relationships, strategic partnership, cloud service or AI integration) that exist between the main platform corporations (*Google, Microsoft, Meta, Amazon, Apple, X, Salesforce*), *microchips* producing companies, *GPU, Tensor Proccessing Unit* (TPU)[5] and those that offer cloud storage services in order to account for the interoperability scenario proposed by this type of technology and its effect on the so-called epistemic bubbles.

Based on this proposal, we aim to map the current status of these technologies through a broad, though not exhaustive or definitive, systematization of

2 See: https://neurips.cc/Conferences/2017

3 IT: digital systems and tools used to process, store, transmit and protect information.

4 DevOps: combines the English words development and operations. It is a set of practices and tools that seeks to improve application development and the release of new software features.

5 TPU – tensor processing unit – it is a type of application-specific integrated circuit (ASIC) developed by Google. It is designed to accelerate machine learning and artificial neural network tasks, specifically optimized to work with Tensor Flow, the open-source library for machine learning. TPUs are known for their high efficiency and performance in processing large volumes of data, making them ideal for AI and deep learning applications.

information. The data systematized in this research was collected from the content of official sites, short company manuscripts, video conferences, specialized news sites, reports and institutional announcements[6]. Through the use of qualitative techniques such as literature review and content analysis, the aim is to build an up-to-date body of knowledge on the current state of the art regarding the integration and development of LLMs. This contribution is especially relevant in a field where the literature remains limited, except for a few key studies. These include works on the genealogy of digital and algorithmic datafication in AI (Gendler, 2024; 2021), the analytical perspective of the different debates surrounding the public conversation on Generative Artificial Intelligences and LLMs (Costa et al., 2023) and studies on *hyperautomation*, understood as the next generation of intelligent automation (Madakam, 2022).

2. ANALYSIS

2.1. LLMs, agency and integration

GPTs are designed to be implemented via APIs[7] that allow developers to integrate their capabilities into various applications and systems. These APIs facilitate communication between the GPT and other programs, allowing text input and receiving responses generated by the model. They are based on a *transformer* architecture that uses natural language processing (NLP) and are pre-trained with large amounts of unsupervised text. These models are then fine-tuned and specialized for specific tasks. Their capacity for continuous learning allows them to expand their reach into multiple areas. Thus, the integration of LLMs into almost any human activity poses new challenges in terms of the ownership and management of the *transformers* that process these data streams. This creates vast control over highly diversified and specific information, the collection of which would be much more difficult without this technology. Striphas's (2015) expression *algorithmic culture* allows us to reflect on this process: "Humans have been delegating the work of culture (…) increasingly to computational processes. Such a change alters the way in which the category of culture has been practiced, experienced and understood until now." (p. 395). Thus, human agency, according to Latour's *actor-network* theory (1999), is made invisible and limited. It sometimes involves "supervision," curation of training data, and parameter settings to adapt

6 Some of the sites consulted were: https://openai.com/about/ - https://www.nvidia.com/es-la/ - https://www.ibm.com/es-es/artificial-intelligence - https://www.salesforce.com/es/company/news-press/press-releases/ - https://ai.meta.com/ - https://deepmind.google/technologies/gemini/ - https://aws.amazon.com/es/ai/ - https://www.anthropic.com/news

7 Langlois et al. (2009) describe the APIs – or application programming interface – as crucial components in the development of Web 2.0, facilitating the emergence of platforms by allowing modular and recursive software interactions. They emphasize the role of APIs in creating structured exchanges between software components, enabling data sharing, automation, and redistribution within the computational culture. Application programming interfaces are sets of protocols and tools that enable communication and data exchange between different software systems. Thus, APIs facilitate integration and interoperability, allowing developers to access the functionalities of other programs without needing to know their internal code).

the model to specific needs. Hence, humans are also responsible for validating the quality of the outputs generated by the GPT and ensuring that they meet ethical and practical standards. But at the same time, such LLMs represent a form of non-human agency in the sense that they can make decisions and perform complex cognitive tasks without direct human intervention: they can analyze data, generate text, perform translations, and answer questions in an automated and efficient manner. Let's take as an example what happens with the integration of LLMs into search engines. Google's indexing algorithm, PageRank, makes decisions based on a set of criteria that evaluate the content of web pages to determine the order in which search results are presented. From these results, the user chooses which hyperlink to click to access the requested information. That is, human and non-human agency working in a simple and collaborative way.

In this equation, the implications related to sponsored links and the monetization system that Google promotes as a business model with *Search Engine Optimization* and *Search Engine Marketing* already weighed[8], although it can be assumed that the subsequent decision is made by humans – mediated by algorithms, but decided by humans–. But what happens when *Gemini* is integrated into the search engine? The answer to that question shapes, in part, the great epistemic dilemmas of our times. First dilemma: the *AI Overview*, the name given to the *click-free* summary response that the search engine provides as the first highlighted answer. This is an automated summary based on criteria that are difficult to determine and track, as they depend on each user's search experience and the LLM's training level at the time of use. In this context, the risk lies in not fully understanding the challenges that AI poses to human agency due to the conceptual ambiguities surrounding the concept of agency. For Floridi (2009), agency refers to the ability of agents (human or artificial) to act autonomously and make informed decisions within the *infosphere*, which is the global information environment. In this sense, agency not only implies the capacity to act, but also the ethical responsibility for the actions taken. A starting point is provided by actor-network theory, which focuses on "collectives" (Latour 2007), not only of humans but also of sets of humans and objects that collectively construct the social and therefore possess agency. A theme that connects with the notion of the *data-driven society*, where the construction of reality is based on the creation of meaning from data, which has driven a crucial debate on the power of agency, including the evolution of human agency with the emergence of an *omnipresent machine agency* (Hepp & Görland, 2024).

8 Search engine optimization, or SEO, is essentially fine-tuning your website to appear naturally or organically in search results on Google, Bing, or any other search engine, and to appear on the Search Engine Results Page (SERP). Search engine marketing—or SEM—is the comprehensive strategy for driving traffic to a site, primarily through paid efforts. Depending on your possibilities and objectives, you can choose the PPC (pay per click) or CPC (cost per click) or CPM (cost per mile impressions) model (Bala & Verma, 2018).

From this theoretical approach arises the second dilemma: the *Google* agency – to continue with the example – becomes central to the information search activity, that is, it becomes an editorial agency. Something that until now was widely debated and refuted by technology companies that do not consider themselves responsible for content generation[9]. This epistemic shift occurs, then, on two levels: 1) *Cognitive,* since it configures new ways of defining reality established and editorialized by non-human agency; 2) *performative,* as it can guide human beings to action in ways that are unimaginable and potentially harmful to the person using it.

Hence, this double game of human and non-human agency enables us to reflect more deeply on the concept of *algorithmic governmentality* of Rouvroy & Berns (2015), a definition that critically observes the functioning of data collection in systems such as those offered by platforms. The authors argue that the main characteristics of this algorithmic government are the creation of a dual reality, a government without a subject, and direct work that limits or extracts the processes of individuation from subjects, or their becoming. Algorithmic governmentality is characterized by a double movement: the abandonment of any form of "scale" in favor of an immanent and evolving normativity in real time, from which emerges a "statistical double" of the world that seems to reject the old hierarchies established by man and the renunciation of any confrontation with individuals whose opportunities for subjectivation are, to say the least, rarefied. The value of this definition lies in its focus on contemporary statistics in relationships and highlights the centrality of algorithmic machines in technological development processes. Costa (2017) highlights the centrality of this notion as a grid of intelligibility of contemporary societies. Something that, in recent years, has become difficult to refute. In this sense, van Dijck (2021) understands that the complexities of platforms are increasingly outdated with respect to the narrow legal and economic concepts on which their governance is based and defines *information ecosystems* as hierarchical and interdependent structures. This inevitably brings into focus issues of power, conflict, and exploitation, a discussion that is currently taking place, for example, with regard to *surveillance capitalism* (Zuboff, 2019) and *data colonialism* (Couldry & Mejías, 2019).

This latter characteristic was greatly enhanced by the partnerships and alliances that occurred between *Microsoft and OpenAI, Nvidia and Meta, Oracle and Microsoft*, to name a few companies, between 2019 and 2024. Examples of this include Microsoft's partnership with *startups* such as *Mistral AI and Alt.ai*, or third parties' use of the infrastructure provided by *Open AI* or *Gemini* (from *Alphabet/Google*) for their development and improvements (Widder, West & Whittaker, 2023). In addition to these movements,

9 Nor does the current legislation in Section 230 of the Communications Act of 1995. On that law, see: https://www.congress.gov/bill/104th-congress/senate-bill/314

the incorporation of AI into major platforms during 2024 demonstrates the enormous interest of corporations in *fine-tuning* according to user needs. Many of these platform integrations don't offer users the option to *opt-out* of their services, such as the *Meta* AI built into *WhatsApp* and *Instagram*. This creates a huge list of countries whose citizens have no way to protect their data and privacy other than by stopping using those services altogether. Table 1 compiles the most recognized and up-to-date integrations of LLMs to various systems, applications and *hardware* in operation. This exercise aims to track the progress made in integrating these systems into the digital ecosystem over the past year.

Tabla 1. Integrating AI into platforms, software, and hardware (2024)[10]

Meta- Meta AI	Meta.ai
	Facebook*
	Instagram*
	WhatsApp (Android-iOS) *
	RayBan Lens (AR)
	Meta Quest (VR)
	You.com
	DuckDuckGo
	APIs
X-Grok	X.com
	Tesla Vehicles
	TeslaBot
	Xai App (iOS-Android)
	You.com
	APIs

10 The data presented in this table is partial and, while intended to be exhaustive, does not represent the entire landscape of integrations, but rather those that most impact the functioning of the platform ecosystem through December 2024.

Google-Gemini	Google.com (AI Overwiews)
	Google Cloud
	Google Chrome
	Google App (Android-iOS)
	Google Notebook
	Google Workspace
	Google Trends
	Google Maps
	Google Lens
	WayMo (Taxi autónomo)
	YouTube
	You.com
	Android (OS)
	Gemini App (Android-iOS)
	Samsung AI*
	APIs
Amazon Bedrock	AWS
	Alexa
	Amazon.com
	Amazon One
	Rufus (Asist. Compras)
	Netflix *
	APIs
Apple Intelligence	Siri
	iCloud***
	Writing Tools***
	Image Playground***

Open AI-ChatGPT	ChatGPT App (Android-iOS)
	ChatGPT.com
	WhatsApp
	Claude (Anthropic)**
	Siri***
	Visual Tools ***
	Compose ***
	Copilot**
	Google Workspace **
	Azure Cloud Services**
	SearchGPT (Beta)
	MercadoLibre (2025) **
	iOS, iPadOS, macOS ***
	DuckDuckGo
	You.com
	APIs
Microsoft-Copilot***	Copilot.microsoft.com
	Microsoft 365
	Bing.com
	Copilot App (Android-iOS)
	Microsoft Teams
	Windows 11
	GitHub
	Hardware Lenovo
	APIs
Anthropic -Claude Sonnet	Claude.ai
	Claude App (Android-iOS)
	DuckDuckGo
	You.com
	Google Docs
	Slack**
	AWS**
	Scribd
	APIs

Perplexity- pplx****	Perplexity.ai
	Perplexity App
	Arc.net
	Uber
	AWS**
	ElevenLabs
	APIs

* Except in the EU (they may refuse to install it).
** Strategic partnership.
*** Open AI LLM Integration.
**** Uses the LLMs from *Open AI, Mistral and Anthropic*.
***** For iOS18, Ipad iOS 18.1 and Sequoia 15.1 (on Iphone 15 Pro and Iphone 16, Ipads and Mac with M1 chip or more).

Source: own elaboration.

2.2. LLMs and concentrated corporate links

Launching text, image, and sound generation tools carries a risk associated with potentially erroneous or malicious results (*deepfakes, fake news, phishing,* among others) that companies seem not to consider. The Western system is largely monopolized by a few technology companies (*Amazon, Alphabet, Meta, Apple, Microsoft*) van Dijck, Poell and de Waal (2018) define a platform as "a programmable architecture designed to organize interactions between users" (p.9) and they classify platforms into two: *infrastructural platforms*, which comprise the main sociotechnical systems for the exchange of resources, messages and content, the coordination of practices, movements and, especially, data and communication flows. Each of them serves the role of providing its infrastructure for a wide range of specific platforms called *sector platforms*—which could not operate without the infrastructure-based foundation. Industry groups increasingly control the gateways to Internet traffic, data flow, and content distribution, and are also able to evade conventional regulatory scrutiny (Gillespie, 2018).

Helmond (2015) defines *platformization* as the penetration of platform extensions into the web and the process by which third parties make their data available to the platform. They interact with APIs, which facilitate the flow of data with third parties (complementers), and *software development packages* (SDKs), which allow third parties to integrate their *software* with the platform infrastructures (Helmond, Nieborg & van der Vlist, 2019). Together, these computing infrastructures and information resources enable institutional relationships that are at the root of a platform's evolution and growth, as platforms provide a technological framework upon which to build (Helmond, 2015). Thus, the transition from the conception of these technologies conceptualized

as platforms to the analysis of their actions as a process, gave rise to the concept of platformization, understood as

> The interpenetration of digital infrastructures, economic processes, and platform governance frameworks across different economic sectors and spheres of life, as well as the reorganization of the cultural practices and imaginaries that exist around these platforms. (Poell, Nieborg & van Dijck, 2022, p. 6)

The explosion of AI assistants applied to word processors, spreadsheets, search engines (*Copilot*, *Gemini*), chat applications such as *WhatsApp*, social networks such as *Instagram* (*Meta AI-Llama*) or *X (Grok)*, and operating systems (*OS*) for various *hardware* such as computers or *smartphones* (for example, *Copilot* or *Samsung AI*), demonstrates how relevant the training provided by individual users is for these companies, thus seeking to improve the performance of these systems. It also forces us to consider the enormous volumes of data these systems process and the deepening of the user profile that this usage entails. ChatGPT 4-0, for example, is designed to recognize the user's environment and incorporate knowledge about their preferences and habits, which are maintained and remembered locally in the GPT's *performance*—what Zuboff (2019) calls *behavioral surplus*. This is a new step in algorithmic governmentality, understood as "a certain type of (a) normative or (a)political rationality that rests on the automated collection, aggregation, and analysis of massive amounts of data so that possible behaviors can be modeled, anticipated, and influenced in advance." (Rouvroy & Berns, 2015, p. 41).

Benjamin Bratton (2016) considers platforms such as smart grids, clouds, and mobile applications that evolve not as separate objects but as a computational apparatus with a new governance architecture layered with diverse interests. The concentration of actors responsible for developing *transformers*, both in terms of programming and training and *hardware* manufacturing, fuels the main concern around the two axes raised in this paper: the deepening of the scope of algorithmic governmentality and the inevitable formation of epistemic bubbles. Currently, there are few companies developing and training *transformers*, and virtually no development by universities or public research centers. Some of them are the *University of Southern California* and *Georgia Tech*. The *National Science Foundation* also made heavy investments in seven newly established National AI Research Institutes across the United States in 2024, but the volumes of money invested are very low compared to the private sector.[11]. In the private universe, the most recognized are *Open AI (GPT),*

11 The announcement of the Stargate project, between Open AI, Oracle, and SoftBank, under the auspices of the Trump administration (January 2025), contemplates collaboration with universities and research centers to promote education in artificial intelligence. This recent initiative appears to seek not only to boost the economy but also to ensure that the U.S. maintains its global technological leadership in the face of significant developments in China.

Microsoft (Copilot); Meta (Llama), Google (Gemini), X AI (Grok) and Apple (Apple Intelligence), all of them in association with the main supplier of *Graphic Processor Units (GPU), Nvidia and Intel*, among others. It is worth clarifying that some companies use open source code, such as *Meta (Llama), Hugging Face (BLOOM), Anthropic (Claude), Cohere (Command), Mistral (Mistral Large), Data Bricks (DBRX) and Perplexity (PPLX*), which also works with an *online* connection to the open *web*. In this context, the Eastern ecosystem, mainly Chinese and Taiwanese, made up of companies such as *01 Ai (Yi Series), H3C (formerly Huawei), Inspur (Yuan), Tencent (Hunyuan), Alibaba (Tongyi Qianwen),* the *Institute of Automation of the Chinese Academy of Sciences* and the *Wuhan Institute of Artificial Intelligence (Zidong Taichu), Zhipu AI (GLM-4), Moonshot AI (Kim), Baichuan, MiniMax (Abab) y Byte Dance* (LLM in development, not yet announced). These companies represent a very competitive and challenging market for the North American market in AI development, as demonstrated by the launch of Deepseek's R1 model in January 2025, which caused million-dollar losses on the stock market for chip-producing companies like NVIDIA due to its enormous potential with lower energy, monetary, and processing requirements.

The LLM universe is a mixed one, combining non-profit organization (NPOs, for its acronym in English) structures with limited liability companies (LLCs, for its acronym in English); others are private for-profit companies (FPOs, for its acronym in English); and some are public benefit corporations (PBCs, for its acronym in English). Some offer their *transformers*, or their models, openly, that is, Open Source (OS), others are closed source (CS, for its acronym in English), while many have such a variety of models that they offer both options to the developer market. For operational reasons, it has not been revealed whether the weights used to pre-train the models are open or closed, as this depends largely on the type of model and the policy of the developing company to share them. As can be seen from this detail of characteristics that this type of organizations can assume, the complexity of the scenario is considerable. Moreover, this diversity is associated with large corporations with a long history and broad recognition. Therefore, it was necessary to reconstruct the links that exist between them in Table 2 in order to observe the links they currently maintain.

Table 2. Investments and partnerships in AI developments linked to LLMs (2024)[12]

LLMs AI StartUps	Platforms								Producers of Cloud Computing Hardware and Services							
	Google	Microsoft	Meta	Amazon	Salesforce	Apple	X	GitHub*	NVIDIA	AMD	Intel	Qualcomm	IBM	Cisco	Oracle	Cerebras
Open AI (ChatGPT) USA- **NPO/LLC CS**		■				■		■							■	
Inflection AI (Pi LLM) USA **FPO-CS**		■						■	■							
Anthropic (Claude) USA **PBC-OS**	■			■	■			■								
Hugging Face (BLOOM) USA **NPO-OS**	■			■	■			■		■	■	■	■	■		
Mistral AI (Mistral Large) FRANCIA **FPO-OS/CS**		■			■			■						■		
Cohere (Command) CANADA **NPO-OS/CS**	■			■	■			■	■					■	■	
Runway ML**(Gen Alpha) USA **FPO-CS**	■				■			■	■							
Aleph Alpha (Luminous) ALEMANIA **FPO-OS**								■					■		■	■
Midjourney* (DaVinci) USA **FPO-CS**	■		■				■	■								
Adept AI (AdetAgent) USA **FPO-OS/CS**		■						■	■							
Character AI (Chat LLM) USA **FPO-CS**	■							■								
Stability AI ***(Diffusion) USA **FPO-OS**	■			■				■	■		■		■			
AI21Labs (WordTune) ISRAEL **FPO-OS/CS**				■				■								
Perplexity AI (pplx-LLM) USA **FPO-OS/CS**	■		■	■				■	■							
Synthesia AI * (Synthesia) INGLATERRA **FPO-CS**	■	■						■	■				■			
TechnologyInnovationInstitute (Falcon) E. ARABES **NPO-OS**				■	■			■								
Eleuther AI (GPT_NeoX) USA **NPO-OS**	■			■				■								
Deepseek (DeepSeek-V2) CHINA **FPO-OS**				■				■	■							
Data Bricks (DBRX) USA **FPO-OS/CS**	■	■	■	■	■			■	■	■	■				■	

* Repository.
** Text to Video, Image to Video and Text to Image.
*** Text to Video, Image to Video and Text to Image- Audio/3D.

Source: own elaboration.

12 The data presented in this table are partial. While intended to be comprehensive, they do not represent the entire landscape of investments and partnerships made as of the date of this article's publication. The original list consisted of 20 StartUps. One of them, MosaicML, was acquired by Databricks in the process of preparing this research, so it was removed from the table. These 20 organizations represent the most advanced developments in Transformers. Note that of the 19, 12 are of American origin, a number that shows the enormous concentration of developments in a single country. According to the 2024 AI Index Report, prepared by Stanford University, the United States leads the list of notable developments in LLMs worldwide with a total of 61, followed by China with 15, France with 8, Germany with 5, and Canada with 4. See: https://hai.stanford.edu/research/ai-index-report

Strategic partnership processes are vital for this new market. Just as Microsoft incorporated *OpenAI* services into its products by offering *hosting* on *Azure*, its cloud, Apple also closed deals in 2024 with *OpenAI* to assist *Siri*, its voice assistant, in its image searches, while *Google* did the same with *Samsung AI* by offering it its *Google Lens* services and *Nvidia* signed commercial agreements for the production of GPUs with *Meta* and *X*, to name a few examples.

3. LLMS, INTEROPERABILITY AND EPISTEMIC BUBBLES

The social and economic costs of concentration in the digital ecosystem constitute a global problem that underpins the economic logic of data extraction that controls the lives of Western consumers (Couldry & Mejías, 2019). This concentration by accumulation—of *hardware*, inputs, traffic, and accumulation of data and training capabilities—entails more challenges than those exclusively linked to excessive accumulation and the processing of social life at levels never seen before. This is not only for commercial and advertising purposes, but also with marked epistemic consequences. Training *transformers* is extremely expensive and consumes very high levels of energy and drinking water. Luccioni (2024) argues that, for the generation of 1000 images, text-to-image AI models such as *DALL-E* require 1.5 liters of water per *kilowatt* of energy consumed. Furthermore, the average CO2 emissions *per capita* are approximately 5 tons per year, while training a large *Transformer* model with neural architecture search emits 284 tons of CO2. Training a single base model (without hyperparameter tuning) on GPUs requires as much energy as a trans-American flight (Strubell, Ganesh & McCallum, 2020). And while some of that energy comes from renewable sources or from the use of carbon offset credits by cloud computing companies, the majority of cloud computing providers' energy does not come from renewable sources, underscoring the need for energy-efficient model architectures and training paradigms.

Furthermore, disruption and innovation in search are not monetary free. The costs of training for an LLM are very high. More importantly, inference costs far outweigh training costs when implementing such a model. In fact, ChatGPT's inference costs exceed training costs on a weekly basis. Dylan Patel, chief analyst at research firm SemiAnalysis.com, revealed in a report published on his website that it costs *OpenAI* 36 cents per query to keep its *chatbot* running until 2024. This means that these companies are aware that the advancement of these models will crush the profits generated by advertising sales, for example, on search engines, in a very short period of time. Hence, they are working against the clock to incorporate AI into all their products in order to generate improvements in their models, reduce latency, and increase user profiling through other means. Whoever integrates AI best and fastest wins. Whoever generates *path dependance* in their users the fastest will be able to implement their products and services most effectively. Some audience insight studies are already observing that the implementation of AI in search engines such as *Gemini* at *Google* or *Copilot* at *Bing* has altered the

number of searches consumers perform, the number of results they click on, or the amount of traffic *Google* sends to the open web or to sponsored links in both the European Union and the United States, affecting its profitability and its business model (Fishkin, 2024).

As can be seen in Table 3, both strategic partnerships, the possibilities of interoperability and the strong dependencies in terms of servers, present a view of extreme concentration in very few actors that should be given special attention if the effect of algorithmic governmentality is to be mitigated.

Table 3. Business-to-business relations in the Western AI market (2024)[13]

Western corporations (BigTech)							
Hardware Producers (H) and Cloud Computing services (CC)	*Google (Gemini)*	*Microsoft (Copilot)*	*Meta (Llama)*	*Amazon (Titan)*	*Apple (Apple Intelligence)*	*X (Grok)*	*Salesforce (X-Gen)*
NVIDIA (H) y (CC) DGX							
AMD (H)							
Intel (H)							
Qualcomm (H)							
Dell (H)							
Cisco (H)							
IBM (CC)							
Oracle (CC)							
Google (CC)	/						
Apple (H) y (CC) iCloud					/		
Microsoft (H) y (CC) Azure		/					
Amazon (CC) AWS				/			

/ Same company
Strategic partnership (financial or hardware supply)
- Cloud computing service
- IA Integration
- Interoperability

Source: own elaboration.

13 The data presented in this table are partial. While intended to be exhaustive, they do not represent the entire landscape of inter-business relations in the Western market, but rather those most closely linked to the platform scenario that is desired to build.

All of the above also deepens the negative effects of *epistemic bubbles*. Following Nguyen (2020), epistemic bubbles form a social epistemic structure in which some relevant voices have been excluded by omission. Epistemic bubbles can form without malicious intent, through ordinary processes of social selection and community formation—which assumes that epistemic cooperation is taking place when subjects communicate with each other—but also through the exclusion of certain groups or individuals, or through the intentional creation of institutional environments that lead individuals to accept one perspective as the only viable one. The latter is what could be in development from the integration of LLMs into the platforms. An effect that causes transmitted, repeated and processed information to generate an amplified system of ideas and beliefs in which concepts or perspectives other than those of the user are omitted (An, Quercia & Crowcroft, 2014). In addition to the editorial agency effect of non-human agents described above, these models work with closed databases or, at best, obtained from the web to generate an *output* that reformulates the content associated with the specific *prompt*. On the *Internet*, English content dominates more than half of all written content online, even though only about 16% of the world's population speaks this language according to the *Internet Society Foundation* (2023). There are more than 7,100 languages in the world according to the *Ethnologue* catalog *(Ethnologue Newsroom,* 2024). Meta's Llama 3.1 works in only 8 languages – at least until July 2024 –, *Gemini* 1.5 *Flash* is available in 40 languages, *Open AI*'s GPT, in its most advanced version, works in 50 languages and in some countries such as China, Russia, Iran and others on the African continent has restricted or limited use. A scenario far removed from the "virtuous circle of AI" heralded by companies like *Nvidia*, whereby any application that uses sequential text, image, or video data is a candidate for *transformers* models.

While researching the properties of language models and how they change with size is of scientific interest, and large LLMs have shown improvements in several tasks, insufficient thought has been given to the potential risks associated with their development and integration into platforms, as well as strategies to mitigate or curb them in a timely manner. It remains to be seen how companies invest, or don't invest, in resources for careful curation and documentation of data sets instead of ingesting everything on the web, reducing the risk of "stochastic parrots" and, therefore, the effects of the epistemic bubble.

> The human tendency to attribute meaning to text, coupled with the ability of large language models to learn patterns associated with various biases and prejudicial attitudes, presents risks of real harm if text generated by these models is disseminated. (Bender et al., 2021, p. 618)

The dangers of synthetic text are closely related to the fact that such text can enter into conversations without any person or entity taking responsibility. They may also have "hallucinations" generating non-existent content presented as real. This responsibility includes both truthfulness and the importance of contextualizing

meaning. Therefore, considering testing these models with significant human intervention in all phases before they are released to the mass market will be very important in order to evaluate how the purpose of the model fits with the research and development objectives and supports the established values (Bender et al., 2021).

4. CONCLUSIONS

The rapidity with which new developments in AI are implemented makes it very difficult for those conducting research in the field to keep up. The information is not transparent; it is often not made explicit in the journalistic material that reports it, and sometimes it is even considered a secret between parties. In addition, the dynamic nature of partnerships, mergers, and intercompany agreements requires a more careful and detailed enforcement framework if we are to understand who controls these systems and how they increase their power. More than ever, there is a need to understand more deeply how platformization works and to map scenarios that contribute to rethinking siloed governance frameworks in a more holistic approach, as suggested by van Dijck (2021).

Integrating LLMs into platforms could expand the scope of algorithmic governance by enabling greater data collection and analysis, which in turn strengthens platforms' ability to model, predict, and respond to user behaviors. When integrated into search engines like *Google (Gemini)* or *Bing (Copilot)*, these language models change the way users access information. The opacity of the criteria used by algorithms to select and present information gives platforms significant editorial agency, raising ethical and epistemic dilemmas. This is augmented by the increasing ability of LLMs to remember and utilize information about users' preferences and habits at the local level, further deepening the concept of behavioral surplus, described by Zuboff (2020). As can be seen in the tables presented, it is vitally important to analyze this context within the framework of platformization, since these technologies will amplify the interpenetration of digital infrastructures, economic processes and governmental frameworks, and the organization of information and online interactions, increasing the concentration of power and control over data and information in the hands of a few. Algorithmic personalization, accelerated by the integration of LLMs into platforms, contributes to the formation of epistemic bubbles. By relying on limited data sets and user preferences, generative language models can deliver results that reinforce preexisting ideas and limit exposure to differing perspectives. Furthermore, the predominance of English in *online* content and the limited availability of LLMs in other languages exacerbate this problem, creating linguistic barriers to accessing information and fostering a diversity of perspectives.

With the rapid integration of AI, and LLMs in particular, into digital platforms, we face significant and urgent challenges. The concentration of power demonstrated by associations and alliances at the tables will lead to a deepening of the

ways in which algorithmic governance is exercised, a problem that requires greater attention from researchers, legislators, and society as a whole. All of this occurs within a poorly regulated framework, which advances at a pace that is difficult to follow and which constantly pushes the boundaries. This, coupled with the lack of transparency in the development and operation of these technologies, as well as the speed with which they are being integrated into various areas of life, hampers the possibility of informed public debate and the implementation of effective control and regulatory measures. The increasing delegation of cognitive processes to AI-related technologies is proving to be a pressing problem in our societies, not only because of the creation of epistemic bubbles but also because it could amplify social, cultural, economic, and political inequalities, while also causing the increased proliferation of racial and discriminatory biases that are difficult to counter with the same speed and power. It will be the task of academia and the researchers involved in these developments to sharpen their critical outlook and dedicate themselves to understanding these phenomena with the greatest possible specificity.

REFERENCES

An, L., Quercia, D. & Crowcroft, J. (2014). Partisan sharing: Facebook evidence and societal consequences. *Proceedings of the Second ACM Conference on Online Social Networks*, 13-24. https://dl.acm.org/doi/abs/10.1145/2660460.2660469

Bala, M. & Verma, D. (2018). A Critical Review of Digital Marketing. *International Journal of Management, IT & Engineering, 8*(10), 321-339. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3545505

Bender, E., Gebru, T., McMillan-Major, A. & Mitchell, M. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Conference on Fairness, Accountability and Transparency,* 610-623. https://doi.org/10.1145/3442188.3445922

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

Costa, F. (2017). Omnes et singulatim en el nuevo orden informacional. Gubernamentalidad algorítmica y vigilancia genética. *Poliética, 5*(1), 40-73. https://doi.org/10.23925/poliética.v5i1.36356

Costa, F., Mónaco, J. A., Covello, A., Novidelsky, I., Zabala, X. & Rodríguez, P. (2023). Desafíos de la Inteligencia Artificial Generativa: Tres escalas y dos enfoques transversales. *Question, 76*(3), e844. https://doi.org/10.24215/16696581e844

Couldry, N. & Mejias, U. (2019). Data colonialism: rethinking big data's relation to the contemporary subject. *Television & New Media, 20*(4), 336-349. https://doi.org/10.1177/1527476418796632

Ethnologue Newsroom (2024). Languages of the World. *Ethnologue*. http://www.ethnologue.com

Fishkin, R. (2024). Zero-Click Search Study: For every 1,000 EU Google Searches, only 374 clicks go to the Open Web. In the US, it's 360. *SparkToro*. https://sparktoro.com/blog/2024-zero-click-search-study-for-every-1000-us-google-searches-only-374-clicks-go-to-the-open-web-in-the-eu-its-360/

Gendler, M. (2024). Datificación social e inteligencia artificial: ¿hacia un nuevo "salto de escala"? *Resonancias*, 17, 121-141. https://doi.org/10.5354/0719-790X.2024.74503

Gendler, M. A. (2021). Mapeando la datificación digital y algorítmica: Genealogía, estado de situación y nuevos desafíos. *InMediaciones de la Comunicación*, *16*(2), 17-33. https://doi.org/10.18861/ic.2021.16.2.3166

Gillespie, T. (2018). *Custodians of the Internet*. Yale University Press.

Helmond, A. (2015). The Platformization of the Web: Making Web Data Platform Ready. *Social Media + Society*, *1*(2). https://doi.org/10.1177/2056305115603080

Helmond, A., Nieborg, D. B. & van der Vlist, F. N. (2019). Facebook's evolution: Development of a platform-as-infrastructure. *Internet Histories*, *3*(2), 123-146. https://doi.org/10.1080/24701475.2019.1593667

Hepp, A. & Görland, S. O. (2024). Agency in a datafied society: an introduction. *Convergence*, *30*(3), 945-955. https://doi.org/10.1177/13548565241254692

Internet Society Foundation (2023). ¿Cuáles son los idiomas más utilizados en Internet? *Fundación Noticias.* https://www.isocfoundation.org/es/2023/09/cuales-son-los-idiomas-mas-utilizados-en-internet/

Langlois, G., McKelvey, F., Elmer, G. & Werbin, K. (2009). Mapping commercial Web 2.0 worlds: Towards a new critical ontogenesis. *Fibreculture*, *14*(2009), 1-14. https://fourteen.fibreculturejournal.org/fcj-095-mapping-commercial-web-2-0-worlds-towards-a-new-critical-ontogenesis/

Latour, B. (2007). *Reassembling the Social: An Introduction to Actor-Network-Theory*. OUP Oxford.

Latour, B. (1999). On Recalling ANT. *The Sociological Review*, *47*(S1), 15-25. https://onlinelibrary.wiley.com/toc/1467954x/1999/47/S1

Luccioni, S., Jernite, Y. & Strubell, E. (2024). Power hungry processing: Watts driving the cost of AI deployment? *The 2024 ACM Conference on Fairness, Accountability and Transparency*. https://dl.acm.org/doi/abs/10.1145/3630106.3658542

Madakam, S., Holmukhe, R. M. & Revulagadda, R. K. (2022). The next generation intelligent automation: hyperautomation. *JISTEM-Journal of Information Systems and Technology Management*, *19*, e202219009. https://doi.org/10.4301/S1807-1775202219009

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, *17*(2), 141-161. https://doi.org/10.1017/epi.2018.32

Poell T., Nieborg D. & van Dijck, J. (2022). Plataformización. *Revista Latinoamericana de Economía y Sociedad Digital*, 1-27. https://doi.org/10.53857/tsfe1722

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

SemiAnalysis (2024). *The Inference Cost Of Search Disruption – Large Language Model Cost Analysis.* https://www.semianalysis.com/p/the-inference-cost-of-search-disruption

Striphas, T. (2015). Algorithmic Culture. *European Journal of Cultural Studies, 18*(4-5). https://doi.org/10.1177/1367549415577392

Strubell, E., Ganesh, A. & McCallum, A. (2020). Energy and policy considerations for modern deep learning research. *Proceedings of the AAAI conference on artificial intelligence*, *34*(9), 13693-13696. https://doi.org/10.1609/aaai.v34i09.7123

van Dijck J., Poell T. & De Waal M. (2018). *The Platform Society*. Oxford University Press.

van Dijck, J. (2021). Seeing the forest for the trees: Visualizing platformization and its governance. *New Media & Society, 23*(9), 2801-2819. https://doi.org/10.1177/1461444820940293

Widder, D., West, S. & Whittaker, M. (2023). Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI. *SRRN*. http://dx.doi.org/10.2139/ssrn.4543807

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. Profile Books.

* Authorship contribution: The conceptualization and comprehensive development of the article was carried out by the author.
* Note: The Academic Committee of the journal approved the publication of the article.
* The dataset supporting the results of this study is not available for public use. The research data will be made available to the reviewers upon request.

IDENTIFICATION OF THE AUTHOR

Agustina Lassi. PhD candidate in Social Sciences, Universidad de Buenos Aires (Argentina). Master in Journalism, Universidad de Buenos Aires. Professor-Researcher, Universidad Nacional de La Matanza (Argentina), Universidad Nacional de Avellaneda (Argentina), Universidad Nacional Guillermo Brown (Argentina) and Universidad Nacional Arturo Jauretche (Argentina). Her research area is sociotechnical studies on digital platforms.