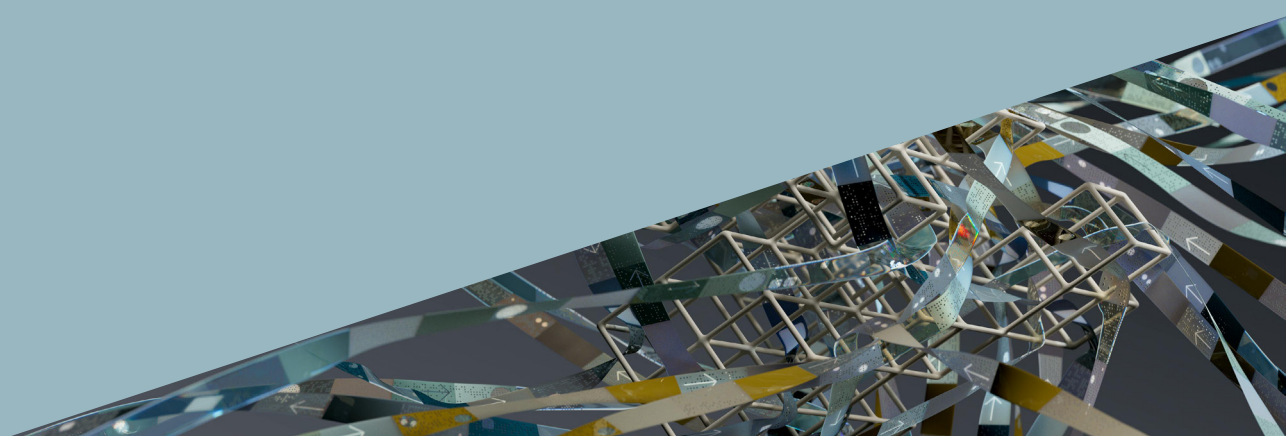
Institut für Auslandsbeziehungen

ifa-Edition Culture and Foreign Policy

Artificial Intelligence and International Cultural Relations

Challenges and Opportunities for Cross-Sector Collaboration

Octavio Kulesz



Institut für
Auslandsbeziehungen

Artificial Intelligence and International Cultural Relations

Challenges and Opportunities for Cross-Sector Collaboration

Octavio Kulesz

Contents

Abbreviations

AI	Artificial intelligence
AISC	AI Safety Institute Consortium
AGI	Artificial general intelligence
API(s)	Application Programming Interface(s)
ARCAI	African Research Center for Artificial Intelligence
ASPI	Australian Strategic Policy Institute
BLOOM	BigScience's Large Open-science Open-access Multilingual Language Model
CEO	Chief Executive Officer
ChatGPT	Chat Generative Pre-Trained Transformer (chatbot)
DACS	Design and Artists Copyright Society
DMA	Digital Markets Act (EU)
DPI	Digital public infrastructure
DSA	Digital Services Act (EU)
EUNIC	EU National Institutes for Culture
FLAIR	First Languages AI Reality (initiative by Mila Institute)
FLOPs	Floating Point Operations per Second (computing performance)
FTC	Federal Trade Commission (US)
G7	Group of 7 countries (Canada, France, Germany, Italy, Japan, the UK and the US, plus the EU)
GAIGI	Global AI Governance Initiative
GDPR	General Data Protection Regulation (EU)
GEMA	*Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte* (collective management association for music in Germany)
GLAM	Galleries, libraries, archives and museums
GPT	Generative Pre-Trained Transformer
ICE	US Immigration and Customs Enforcement
ICR	International Cultural Relations
IF	Institut Français
IMF	International Monetary Fund
IoT	Internet of Things
IP	Intellectual property
IPPR	Institute for Public Policy Research
IT	Information technology

LEIA	*Lengua Española e Inteligencia Artificial* (project)
Llama	Large Language Model Meta AI (model from Meta AI)
LLM(s)	Large Language Model(s)
NLLB	No Language Left Behind (an initiative by Meta)
NLP	Natural Language Processing
NPC	Non-player characters (in games)
OECD	Organisation for Economic Co-operation and Development
R&D	Research and development
RLHF	Reinforced learning with human feedback
SACEM	*Société des Auteurs, Compositeurs et Éditeurs de Musique* (collective management organisation for music in France)
SGE	Search Generative Experience
SLM(s)	Small Language Model(s)
STT	Speech-to-Text
TB	Terabyte(s)
TFGBV	Technology-facilitated gender-based violence
TTS	Text-to-Speech
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UK	United Kingdom
US	United States (of America)

Foreword

The development and integration of AI systems forms a part of the digital revolution. In its capacities to transform social interactions, however, it surpasses any of the previous digital tools. As AI will impact communication as well as education, arts and politics, it will also impact international cultural relations.

The author outlines opportunities for the art sector like AI's capacities for an enhancement of creative potentials. Likewise, translation tools might ease international collaboration. However, the author focusses on the challenges involved to domestically and internationally safeguard human rights and equality in the AI environment.

International cultural relations are challenged by AI on multiple dimensions: cybersecurity and data protection risks to democracy; Global North and South divides in development and protection measures that undermine international trust; a geopolitical undercurrent in the technological race of global superpowers; a power concentration on private companies that require tech diplomacy; specifically, cultural diversity and exchanges are challenged by non-inclusive content selection systems that might deepen further global divides. Moreover, a minority's mighty reductionist technological worldview might need powerful cultural refutations.

As the author states, the human dimension of this technological shift needs the broad attention of international cultural relations actors: "Cultivate intercultural understanding as a foundational element, recognising that a technological system cannot adequately guarantee peace, safety or fairness by itself."

This study forms part of ifa's Research Programme "Culture and Foreign Policy", in which experts address relevant issues relating to culture and foreign policy with the aim of involving academics, practitioners, policymakers and the civil society. I would like to thank Octavio Kulesz for his excellent work and commitment to this research. In addition, I would like to thank my ifa colleagues Sarah Widmaier and Ivana Putri for their work on the conception, coordination and editing of this project.

The publication is right on time. In September 2024, the UN Summit of the Future will release *The Global Digital Compact* that is expected to "outline

shared principles for an open, free and secure digital future for all"[1]. It is about time to raise awareness and discuss, to construct knowledge landscapes and informed publics. International cultural relations will have to skilfully integrate this technology in its practices just as much as it will need to use its competences to safeguard democracy and human rights in this fast-evolving technopolitical complex.

Odila Triebel
Head of Dialogue and Research
ifa – Institut für Auslandsbeziehungen

[1] https://www.un.org/techenvoy/global-digital-compact.

Abstract

The integration of AI systems has profoundly transformed multiple dimensions of international cultural relations (ICR). Notably, changes in recent years have accentuated imbalances among languages, affected the situation of cultural sectors, and impacted the integrity of information while also providing a platform for influential tech players with unique imaginaries. This is unfolding in an era characterised by an increasing geopolitical competition and a global search for new AI governance frameworks. Most of these issues will likely further intensify in the coming years, with the proliferation of AI agents with whom users will interact in various contexts. In such a scenario, it has become urgent to outline a strategy for the work in the field of ICR in the AI era. Considering the mutual influence between AI and ICR, institutions engaged in ICR should aim to ensure ethical and responsible technology use and emerge as pivotal contributors to the advancement and regulation of AI.

Executive Summary

The integration of AI systems presents both opportunities and challenges across the key domains of international cultural relations (ICR). Indeed, virtually no aspects of ICR remain unaffected by this technology, given that it is so deeply entwined with cultural dynamics.

From a technological standpoint, AI systems allow for the execution of an increasing number of complex tasks. However, they may suffer from issues such as the so-called "hallucinations", the potential for a model collapse due to a lack of original content, embedded biases, value conflicts and security vulnerabilities, among other problems.

When it comes to languages, although AI systems can contribute to areas such as translation, language learning and the preservation of linguistic heritage, enormous imbalances between languages—particularly between English and all others—exist as well as challenges that could threaten the very survival of low-resource languages.

In the creative sectors, AI has an undeniable impact. Integrating this technology into literature, visual arts, cinema, museums, video games, to name just some, offers expanded artistic possibilities, enhanced productivity for cultural and creative industries and a significantly enriched content range. However, the lack of transparency in using artworks to train AI systems, combined with the potential displacement of artists and entrepreneurs along the cultural value chain, could lead to a rapid degradation of the creative ecosystem.

In the information environment, AI tools can potentially improve communication and streamline content moderation. Nonetheless, here, challenges are escalating daily with a proliferation of synthetic content, misinformation, disinformation and AI-driven manipulation. Simultaneously, the strategies employed by search engines, social networks and other platforms can lead to a further decline of the situation. Additionally, censorship methods in the age of AI are becoming more sophisticated.

Regarding technological agents, although new AI startups emerge monthly, the few that survive eventually align with major tech corporations in a sector marked by a conspicuous lack of diversity. In this landscape, risks such as economic concentration, job loss and growing inequality are expected to intensify in the

future. Furthermore, the ideologies of tech leaders often take on rather striking forms, while the professed commitments of many of these individuals frequently contradict their actual practices.

Finally, international relations in the AI era are not free from significant geopolitical tensions, particularly due to an escalating rivalry between the US and China. Globally, AI strategies and regulations are proliferating, though currently, relatively few states have managed to design and adopt long-term policies. Especially countries in the Global South are at risk of becoming subjected to a colonial-like exploitation of data and resources. Despite these challenges, several global frameworks have emerged under the auspices of entities such as UNESCO and the broader UN system, along with a growing number of tech diplomacy initiatives that contribute to a closer engagement among the various stakeholders.

In this scenario, the work for ICR institutions can be challenging. As expressed by the protagonists themselves, more data, skills and collaborations with the tech sector are necessary. Simultaneously, the realm of imaginaries and narratives around AI represents an enormous untapped potential when it comes to influencing the development and regulation of AI. Several projects have been initiated, intersecting ICR and AI, though much remains to be done, especially when it comes to the promotion of a new approach that could be described as "Digital Humanism."

In this context, the study presents a series of recommendations aimed at actors working directly or indirectly in the field of ICR, with a particular focus on ICR institutions.

The initial recommendations focus on enhancing AI and data literacy through training sessions that not only cover the technological aspects of AI but also address its social implications. This education targets a broad range of stakeholders, from government representatives to the civil society at large, emphasising lifelong learning and the development of essential soft skills like empathy and cross-cultural communication.

Further suggestions push for stronger links among ICR institutions globally to facilitate the sharing of best practices and responsible AI usage. This involves

fostering dialogues with tech actors and engaging various AI subcultures to explore AI's potential in creative industries and ensure the inclusion of diverse perspectives, especially from underrepresented communities.

Concerning the policy, it is recommended that AI research and policy-making processes include a wider array of voices beyond tech companies and government bodies. Emphasizing cultural aspects in national AI strategies and encouraging active participation in international forums are a key for the promotion of cultural diversity and the protection of cultural heritage. Furthermore, new debates around AI ethics and governance should be sparked, advocating for discussions challenging existing AI narratives. This includes probing the societal impacts of AI on concepts like "truth" and "shared reality".

For AI development and deployment, a diversified approach would be best, supporting specific research communities and pushing for transparency across all AI operations. Promoting ethically sourced and open datasets is crucial to the maintaining of trust and integrity in AI applications.

Lastly, the relationship between AI and ICR is portrayed as a dynamic, reciprocal interaction in need of adaptable strategies. Promoting new roles like "digital humanist" and continuous reassessments of AI strategies are suggested to keep pace with technological advancements.

Introduction

Over the past five years, the integration of Artificial Intelligence (AI) systems has profoundly and globally transformed multiple dimensions of social life. Particularly noteworthy is the widespread adoption of consumer-oriented tools such as ChatGPT, which can be used without technical knowledge and which have significantly influenced fields as varied as education, communications, the arts, commerce and politics, among others. AI is another step in a long list of prior digital evolutions, like social media and mobile phones, that have reshaped social interactions. However, the pace and magnitude of the current shift far exceeds any previous technology. Each month, new models and new market participants introduce another layer of complexity to the current scenario.

International cultural relations (ICR) have not been immune to this extensive transformation. Understanding ICR as "reciprocal transnational interactions between two or more cultures, encompassing a range of activities conducted by state and/or non-state actors within the space of culture and civil society" (British Council and Goethe-Institut, 2018: 6), it is challenging to find an aspect of ICR unaltered by the ongoing revolution. The expanding presence of AI in the cultural sectors and the global information ecosystem is just one of many indicators of an unprecedented shift in ICR.

Future developments promise to be even more impactful. Currently, leading AI systems enable users to perform isolated tasks such as creating text or images, but a future scenario in which interactions become more intricate—not only human-AI but also AI-to-AI and all possible subsequent combinations thereof, such as human-AI-human or AI-human-AI—could propel ICR into uncharted waters.

As with any technology, AI possesses a dual nature, offering ample opportunities while posing significant challenges. From striking a balance between capitalising on these advantages and mitigating the associated risks, insights can be gained, contributing to a more ethical and democratic utilisation of AI systems. Institutions engaged in ICR undoubtedly have a crucial role to play in this endeavour.

However, beyond merely monitoring the proper use of tools, participants in ICR can also assume an innovative role during this pivotal moment for a technology such as AI, which is closely intertwined with human culture. This

new responsibility demands proactively reshaping the imaginaries and narratives that fundamentally drive the development and regulation of AI systems on an international scale.

Indeed—as will be explored in subsequent chapters—, many concepts central to AI research, such as the notion of data as something objectively "given," are deeply rooted in cultural assumptions. At the same time, the names of numerous AI applications and commercial brands—like Grok and DeepMind—are inspired by science-fiction literature. Furthermore, the ideologies underlying the discourses of tech leaders often exhibit idiosyncratic characteristics. Additionally, AI regulatory frameworks across different world regions are not uniform but shaped by local political, legal and cultural realities. All these elements, embedded in various aspects of AI, are subsequently projected across different territories, forming a complex landscape of sociotechnical interrelations.

Ultimately, it is crucial to recognise that the link between AI and ICR is not unidirectional. Not only does the integration of AI systems influence the various facets of intercultural relations, but there is also a reciprocal effect: the development and regulation of AI are often driven by specific imaginaries and practices that are imbued with a strong cultural component, upon which it is possible to act.

Addressing these issues becomes urgent in a context of increasing geopolitical tensions, where the most powerful nations compete in an "AI race" with uncertain outcomes. Simultaneously, the rise of tech giants who relentlessly extract data for their AI systems and often disregard consequences such as the erosion of public discourse, the disruption of cultural and informational ecosystems, the loss of cultural and linguistic diversity, and irreversible environmental degradation, provides another reason for swift actions.

In this context, the voice of institutions working in ICR can attempt to pivot competition towards cooperation and mutual understanding, collaborating with a diverse array of stakeholders from both public and private sectors, as well as civil society, from the Global North and the Global South. These partnerships may be vital in the process and necessitate the adoption of new types of participatory and bottom-up methodologies.

In this sense, this report does not aim to provide merely a list of AI tools applicable in ICR. Such an endeavour would be futile given the rapid evolution of automation technologies, while it would also neglect the broader, more qualitative long-term impacts on both sides and in both directions. Instead, this report seeks to delineate the key technologies, stakeholders, processes, narratives and tensions across various points of intersection, in an effort to foresee medium and long-term trends and to contemplate strategic actions.

It must be acknowledged that the topic at hand encompasses multiple complexities. Firstly, the volume of information available is enormous, updating at high speed. Moreover, these issues present both quantitative and qualitative aspects, thus demanding a multidimensional approach—that is, one that incorporates, among others, technological, sociological, historical and artistic considerations. Furthermore, it is a global issue, without allowing for a uniform approach, as local manifestations are of extremely diverse nature.

Therefore, this research draws from multiple sources of information. It is based on an extensive array of scientific articles, press notes and book chapters, most of which were published after 2022. Additionally, to understand the opportunities and challenges of ICR in the AI era firsthand, interviews with 14 experts from various countries, both from the North and the South, were conducted.[2]

To structure the information, this report is divided into seven chapters. Chapter 1 offers an *overview of AI systems*, their applications and technical issues. Chapter 2 addresses the *complex interplay between AI and languages*. Chapter 3 focuses on the impact of AI on *creative sectors*. Chapter 4 examines various aspects related to the *use of AI systems in the information sphere*. Chapter 5 conducts an *analysis of big tech companies*, particularly those based in Silicon Valley, while Chapter 6 explores the *geopolitical dimensions*, including AI strategies and regulatory frameworks. Chapter 7 specifically considers the *role of institutions working in ICR* in addressing emerging challenges. The report finishes with a set of concrete recommendations to these organisations.

[2] These interviews took place from January to April 2024. See the complete list of interviewees in the Appendix.

1. AI Systems: Features, Models, and Challenges

1.1 Definitions

The definition of Artificial Intelligence (AI) varies considerably among different institutions and scholars. However, it is important to note that most definitions, whether crafted by governments or private companies, characterise AI as a collection of operations and processes, deliberately avoiding any anthropomorphic views of the technology. For instance, the United Nations Educational, Scientific and Cultural Organisation (UNESCO) defines AI systems as those possessing the "capacity to process data and information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning, or control" (UNESCO, 2021: 10). Google, on the other hand, describes AI as "a set of technologies that enable computers to perform a variety of advanced functions, including the ability to see, understand and translate spoken and written language, analyse data, make recommendations, and more" (Google, 2022).

1.2 The Current AI Boom

Since its formal inception as a field of study in the 1950s, AI has experienced various phases of rapid growth and stagnation, often referred to as "AI summers" and "AI winters". Until the beginning of the 21st century, AI development was primarily concerned with creating automated solutions that operated based on predefined rules. The focus of AI research has dramatically evolved since the 2010s towards machine learning—a technique that allows machines to learn from large datasets. Special emphasis is placed on deep learning, a specialised branch of machine learning that models high-level abstractions in data through the use of multiple processing layers comprising artificial neurons, drawing inspirations from the structure of the human brain.

The current AI boom, the most significant in history, can be attributed to several factors. Firstly, the digitisation of all aspects of social life and the exponential growth of user-generated content over the last two decades have provided access to unprecedented amounts of data. Furthermore, AI systems, particularly those based on deep learning, depend on sophisticated models, many of which were developed in the past five years. Another key factor is the growing pool of talent

in the field. Data science, once a niche profession, has now become one of the highest-paying jobs globally. Additionally, successive improvements in computing power and increasing investments in infrastructure have significantly contributed to AI's expansion.

1.3 The Generative Era

From a cultural perspective, AI has been a focal point of study and intense debate for many years, particularly concerning the impacts of recommendation algorithms. Indeed, online-content selection systems present significant challenges for cultural diversity and cross-cultural exchanges. This is largely due to the opaque nature of selection criteria on major platforms, as well as the emergence of phenomena such as filter bubbles—or echo chambers—, which can limit the exposure to diverse viewpoints (Pariser, 2011). However, the advent of generative AI, which enables the automatic creation of various types of content, represents a qualitative leap that introduces opportunities and challenges on an entirely different scale. In what follows, the research will explore some of the most recent AI tools that are profoundly transforming the landscape of ICR.

One of the most remarkable advancements in AI is the development of Large Language Models (LLMs). These sophisticated AI systems are trained on vast amounts of text data with the primary function of statistically predicting the next word in a sequence. This capability allows LLMs to generate coherent and contextually relevant text based on the given input or prompt. Beyond simple word prediction, LLMs are employed in a wide range of Natural Language Processing (NLP) operations, such as text generation, summarisation, translation, keyword extraction, sentiment analysis and even coding. The efficacy of these models increases as they are fed larger and more refined datasets (Kaplan et al., 2020). LLMs form the foundation of modern chatbots like ChatGPT (OpenAI), Claude (Anthropic) or Pi (Inflection), among many others.

The infrastructure needed to support LLMs may involve dozens or hundreds of thousands of servers, leading to a considerable energy expenditure. In contrast, Small Language Models (SLMs) are emerging as a promising alternative, offering

several advantages such as faster training times, lower latency and improved energy efficiency (Taulli, 2024).

In addition to LLMs, other groundbreaking AI technologies are transforming the cultural landscape. Diffusion models for image creation, such as Midjourney, Stable Diffusion (Stability AI) and Dall-E (OpenAI), have taken the world by storm, enabling users to generate stunning visual artwork from textual descriptions. Video generation is another frontier in AI creativity, with models like Sora (OpenAI) showcasing the potential for creating dynamic, animated content from scratch. Multimodality is becoming a key characteristic of current AI systems, with LLMs increasingly being integrated with Text-to-Speech (TTS) and Speech-to-Text (STT) technologies to handle both written and spoken language. Moreover, tools like ChatGPT not only allow for the generation of images from text but also enable the creation of text from images, further pushing the boundaries of multimodal interaction.

It is worth noting that these AI systems are not only used by everyday users but also frequently accessed via Application Programming Interfaces (APIs). APIs allow companies to seamlessly integrate these powerful AI capabilities into their own websites, enabling real-time data processing at scale.

AI systems can be classified as either open or closed. Open systems have their source code and sometimes even the trained models publicly available, allowing for anyone to access and modify it, such as Stable Diffusion. Closed or proprietary AI systems keep their source code and operational details confidential, accessible only to authorised personnel, such as Midjourney or ChatGPT.[3]

In recent years, new expressions have gained popularity to refer to these cutting-edge AI systems: "general-purpose AI" (AI systems that can perform tasks across multiple domains), "foundation models" (models trained on broad data that can be adapted to a wide range of applications) and "frontier AI" (highly capable general-purpose AI models that can perform a variety of tasks and match or exceed the capabilities present in today's most advanced models).

[3] Ironically, despite the "open" in OpenAI, most of its tools are not openly accessible.

1.4 AI Agents, Human Interactions and the Metaverse

All these developments are paving the way for a new generation of AI systems known as "advanced AI assistants" or "AI agents". These agents are significantly more complex than the current LLMs. They are designed to receive inputs in various formats—text, audio, video—and are equipped with memory and the capability to act based on specific goals or objectives. Additionally, they can interact with other agents. For example, while an LLM might just provide information about places to visit, an AI agent could take further actions such as finding the best hotel, making reservations and communicating with the hotel about the guest's arrival time. These agents may also have self-improving capabilities.

In March 2024, a start-up named Cognition AI showcased a programme called Devin performing tasks typically done by software engineers. Devin not only generated code—a task achievable by systems like ChatGPT or Claude—but also went a step further by devising solutions to technical problems, first writing the code, then testing and implementing it (Vance, 2024).

AI agents will be able to extend their capabilities beyond online operations and will be interconnected with objects in both peaceful and military applications. This encompasses industrial robots, robots that assist or interact with humans, transportation systems and home appliances, all integrated under the broader umbrella of the Internet of Things (IoT).

The increased efficiency of LLMs and other AI systems will facilitate and accelerate the communication between humans and AI agents, between agents, and in even more complex interactions such as human-agent-human, leading to the development of a widespread ecosystem of agents. Social networks are evolving to include AI characters with whom users will interact in increasingly synthetic environments, signalling a shift toward more immersive digital experiences. OpenAI has explicitly indicated that scaling video generation models such as Sora could be a promising avenue toward creating general-purpose simulators of the physical world (OpenAI, 2024). Although it is still early days, the convergence of AI and the metaverse—a collective virtual shared space that merges enhanced physical realities with persistent virtual environments—is poised to bring profound changes to human interactions.

1.5 Transparency, Explainability and Accountability

It is clear that AI systems have significant social impacts, necessitating ethical reflection. At this juncture, it is useful to introduce three crucial attributes: transparency, explainability and accountability. *Transparency* ensures that the operations of AI systems are open and observable, allowing users to understand the decision-making process. *Explainability* involves the capability of AI systems to provide understandable explanations for their decisions and actions to users. *Accountability* means that there must be individuals or entities responsible for the actions and decisions made by an AI system.

In any case, it makes no sense to blame the AI system itself for possible adverse outcomes, as if it possessed the capacity for conscious decision-making and self-awareness. Indeed, inquiring about the consciousness of an AI system is nearly tantamount to seeking consciousness in an Excel formula (Brodsky, 2024). Thus, when things go wrong, it is essential to *"chercher l'humanité"* (i.e. seek the human component), as noted by Floridi (2023). An illustrative example is the case of Air Canada defending itself in a tribunal by claiming it could not be held accountable for information provided by its chatbot. The tribunal ruled that it was a significant oversight to suggest that the chatbot was a separate legal entity responsible for its own actions. As Tribunal member Christopher C. Rivers pointed out,

> *"While a chatbot has an interactive component, it is still just a part of Air Canada's website. It should be obvious to Air Canada that it is responsible for all the information on its website. It makes no difference whether the information comes from a static page or a chatbot."* (as quoted in Yagoda, 2024)

1.6 Hallucinations and a Lack of Understanding

At this point, it is essential to acknowledge that current AI systems exhibit numerous weaknesses that can often lead to unintended consequences. Many of these shortcomings were already anticipated by researchers in the early years of

large language models (LLMs), such as Bender, Gebru and other experts, who characterised LLMs as "stochastic parrots" (Bender et al., 2021). One key challenge associated with these models is "hallucinations", where the AI generates information not present in its training data or that is outright false, often due to its inability to truly understand context or verify facts (Weise and Metz, 2023).

When discussing language use, Bender and Koller (2020) emphasise that human communication serves a purpose; it is not merely about the act of speaking but about achieving specific communicative intents, such as informing, requesting or socialising. However, LLMs do not grasp the meaning behind words; they merely learn the distribution of language patterns, which results in a limited capacity to understand nuances like irony or humour.

This issue has sparked intense debates at all levels, especially when some scientific circles assert that the human brain is nothing more than a machine—and conversely, that machines could one day think like humans (Cobb, 2020). Even Pope Francis has contributed to the discussion by taking a clear stance on the matter:

> *"No doubt, machines possess a limitlessly greater capacity than human beings for storing and correlating data, but human beings alone are capable of making sense of that data."* (Holy See Press Office, 2024)

1.7 Model Collapse

Additionally, it is important to note that there is currently an exponential growth in synthetic content, while new content created by humans is becoming scarcer—to an extent that high-quality text data could be gone in a few years (Villalobos et al., 2022). Should this trend continue, the quality of AI outputs may dramatically deteriorate.

Indeed, when an AI system is increasingly fed synthetic data instead of content originally created by humans, a phenomenon known as "model collapse" can occur (Shumailov et al., 2023). This refers to a situation in which the AI begins

to degenerate, losing its ability to generalise from its training, and producing repetitive or highly similar outputs. This deterioration can lead to more pronounced errors in the content subsequently generated, differentiating it from the typical inaccuracies or hallucinations seen in AI outputs.

1.8 Biases, Value Conflict and Intellectual Property Infringement

Furthermore, far from being neutral tools, AI systems are embedded with biases and stereotypes related to gender, ethnic origin, religion, politics and other factors. For instance, a recent report by UNESCO-IRCAI highlighted that LLMs are more likely to associate female names with "home," "family" and "children", while it does the same for male names with "business," "executive," "salary" and "career." In scenarios in which LLMs were prompted to complete sentences that began with a mention of a person's gender, sexist and misogynistic content was produced in about 20 % of cases, including disturbing stereotypes about women (UNESCO-IRCAI, 2024: 3).

In image generation systems, the problem of biases is evident in a study by *Rest of World*—a nonprofit publication focusing on technology experiences outside the Western perspective—, which analysed 3,000 AI-generated images using Midjourney. This study exposed how these systems often stereotype national and gender identities. For example, images of "an Indian person" predominantly depicted elderly men in traditional attire, while "an American person" frequently showed young, light-skinned women, reflecting the US-media biases (Turk, 2023).

One of the most common causes of biases in AI systems relates to the data used for training. Many current LLMs are trained on datasets compiled from indiscriminate sources, such as Common Crawl, which aggregates vast amounts of web data and does not filter out hate speech content. Using such unfiltered data for training LLMs can lead to models that produce undesirable or harmful outputs. Moreover, Common Crawl does not capture the "entire web", which further complicates the quality and representativeness of the training data (Baack, 2024).

The issues associated with biases and stereotypes are not isolated problems. Instead, it can be asserted that—due to the types of data used and the processes employed—LLMs may propagate a comprehensive worldview. Indeed, some of the major current LLMs tend to reflect many values dominant in the US on topics such as immigration, gender and freedom of speech, which may conflict with the values of other countries or regions (Johnson et al., 2022).

Additionally, the indiscriminate use of data raises serious intellectual property (IP) concerns. Training AI systems with data from diverse sources can infringe on copyright and proprietary rights, posing legal and ethical challenges, particularly in the arts, in cultural and creative industries where the unauthorised use of copyrighted material can lead to significant conflicts.[4]

1.9 AI Alignment

The issue of biases in AI systems has spurred the development of an entire research field known as "AI alignment", focusing on ensuring that AI behaviours align with human values and ethical standards. One initiative in this area is the concept of "Constitutional AI" proposed by the company Anthropic, which uses a set of guiding principles—akin to a constitution—to steer AI outputs. This constitution is designed to encourage normative behaviours that prevent toxic or discriminatory outputs. Nevertheless, according to Anthropic, many of their principles were developed through "a process of trial and error" (Anthropic, 2023), raising questions about the extent to which these can be considered true "principles". This iterative approach also prompts scrutiny regarding the values of those involved in the process.

Another commonly employed approach in AI alignment is Reinforced Learning with Human Feedback (RLHF). This involves training AI systems by reinforcing desired outputs through feedback provided by human overseers. Such an approach can significantly enhance the reliability of AI outputs. However, this method is not without its limitations regarding biases: also here, a critical aspect is the values of the individuals providing human feedback, as well

[4] This topic will be addressed in more detail in Chapter 3.

as the criteria in the selection of said individuals. Therefore, while reinforced learning and other techniques can significantly enhance the performance of AI systems, they do not eradicate biases, particularly those introduced by trainers.

Furthermore, no technical process will eliminate the blatant stereotypes expressed when AI-based services are introduced to the market. For example, the default assignment of female names and voices to major voice assistants like Alexa, Cortana and Siri reflects not a technical necessity but rather developers' biases (West et al., 2019).

Additionally, it is worth noting that many of the basic concepts in the typical vocabulary of the AI world might be considered problematic in themselves. For instance, the term "data"—derived from the Latin "datum," meaning "that which is given"—does not adequately reflect the complexity of its use. Data is not a neutral entity that presents itself on its own but is shaped by interpretation and manipulation, stripped of context for broad comprehensibility, and categorised through politically and socially influenced systems that reflect selective interests and inherent biases (Thi Nguyen, 2024). Therefore, it would be far more appropriate to use the term "capta," meaning "that which is taken," instead of "data" (Drucker, 2011).

In any case, the concept of "alignment"—which suggests a somewhat geometric approach to ethical issues—does not seem entirely adequate for analysing or effectively addressing the social impact of AI systems. After all, there is never a value-neutral standpoint from which to analyse human language and culture (Weatherby, 2023).

1.10 AI Safety: Vulnerabilities, Privacy Risks and the Escalation of Cybersecurity Threats

AI safety has emerged as an extensively studied field in recent years. It goes beyond analysing the ethical impact of AI systems in terms of biases and discrimination to consider the vulnerabilities and risks inherent to these technologies. Since 2023, significant resources have been allocated to scientific

research and high-level discussions on this issue, which has also garnered a hype in the media.

Besides the extreme catastrophic risk of a supposed "rogue AI" destroying the world—a topic heavily influenced by Silicon Valley leadership imaginaries that will be detailed in Chapter 5—, AI systems do raise numerous questions in terms of safety and stability. For example, vulnerabilities such as *jailbreak attacks* allow malicious users to manipulate AI prompts to bypass intended usage policies. *Prompt injection* techniques enable the insertion of specific inputs into AI systems to deliberately manipulate their outputs or behaviour, while *data poisoning* involves corrupting the training data of an AI model to adversely influence its future responses.

Safety perspectives vary between supporters of open and closed AI models. Many contend that open-source systems are less secure because they allow any actor, including those with malicious intent, to access highly effective models to then use them to cause massive damage (Harris, 2024). Conversely, proponents of open systems argue that their inherent transparency actually enhances security: since an open system is reviewed by a much larger community, there is a higher probability of identifying and addressing security threats, whereas closed systems rely on a restricted access to maintain security (Open for Good Alliance, 2023).

As some of the aforementioned vulnerabilities can expose sensitive information, they highlight the ongoing challenges in protecting personal data in AI systems. Moreover, the widespread use of these technologies concurrently ushers in an era of unprecedented surveillance which poses threats to privacy, as noted by Harvard experts Bruce Schneier and Nathan E. Sanders concerning AI services by major tech players:

> *"If searching on Google in the 2010s was like being watched on a security camera, then using AI in the late 2020s will be like having a butler. You will willingly include them in every conversation you have, everything you write, every item you shop for, every want, every fear, everything."* (Schneier and Sanders, 2023)

It is also important to note that AI systems can be used to carry out increasingly complex and automated cyberattacks. Once AI-driven malware enters a system, it is capable of mutating into multiple forms to evade detection (Schmidt, 2022: 293). This adaptability makes AI-powered threats particularly difficult to counteract.

1.11 A Staggering Environmental Impact

Another significant issue with current AI systems is their environmental impact, particularly in terms of electricity, water and other natural-resources consumption (Crawford, 2021). Researchers at the AI platform Hugging Face and Carnegie Mellon University have highlighted that generating an image using a powerful AI model can consume as much energy as fully charging a smartphone (Luccioni et al., 2023). They also found that large models are far more energy-intensive than smaller, task-specific models. For instance, classifying movie reviews as positive or negative using a large model consumes about 30 times as much energy as using a model finely tuned specifically to that task.

Additionally, a 2023 report has raised concerns that Google's AI operations could soon use as much electricity as the whole country of Ireland (De Vries, 2023). The escalating energy demand suggests that even with efforts to enhance data centre energy efficiency, the growth of the AI sector could lead to an 80 % increase in its emissions that contribute to global warming, should the number of data centres double to meet industry needs (Climate Action, 2024).

2. AI, Languages and Diversity

As with languages, AI systems offer significant opportunities. The use of multilingual AI-based tools can better the translation of languages and a real-time interpretation, facilitating a more effective communication among individuals from diverse linguistic backgrounds. These capabilities might also extend to automated subtitling and dubbing, making content more widely accessible. Furthermore, AI systems could support language learning through interactive and adaptive platforms that cater to the diverse needs of learners, potentially accelerating language acquisition.

AI systems also play a potential role in the preservation of endangered languages by enabling the documentation and revitalisation of linguistic heritages that might otherwise face extinction.

However, while the integration of AI technologies offers new possibilities, it also presents significant challenges such as imbalances among languages and technical hurdles, which are further elaborated below.

2.1 English as the Dominant Language

A critical issue is the overwhelming dominance of the English language in the digital sphere. Despite ranking third in terms of first-language speakers and being spoken by less than 20 % of the world's population, more than half of all existing websites are in English, establishing it as the primary language on the internet (Richter, 2024).

As Gabriel Nicholas and Aliya Bhatia explain in their comprehensive report, *Lost in Translation: Large Language Models in Non-English Content Analysis*, the dominance of the English language does not stem from any inherent linguistic superiority but is largely, among other factors, due to the historical legacies of British colonialism and the significant influence of American companies in the global economy (Nicholas and Bhatia, 2023: 9). The pre-eminence of English extends to the sciences, particularly in NLP, where it serves as the primary language for scientific communication and the main language for datasets used to train AI systems.

2.2 High-Resource vs. Low-Resource Languages

At this point, it is worth introducing the distinction between "high-resource" and "low-resource" languages. English is classified as an *extremely high-resource* language, benefiting from a wealth of high-quality data resources. Other languages such as Chinese, Spanish, French and Japanese are still considered *high-resource* languages despite the availability of fewer data resources compared to English. There are also several dozen *medium-resource* languages, such as Italian and Urdu. However, the majority of the world's over 6,000 languages are categorised as *low-resource* or *extremely low-resource*, with very limited text available. The data quality for these languages is often compromised, suffering from issues such as mistranslations, nonsensical texts scraped from the internet, or content confined to narrow domains like religious texts and Wikipedia.

Nicholas and Bhatia (2023: 15) point out that the disparity in data availability, known as the "resourcedness gap", highlights a disconnect between the availability of language resources and a language's actual number of speakers or internet users. For instance, despite having hundreds of millions of speakers, languages like Hindi, Bangla (Bengali) and Bahasa Indonesia (Indonesian) are still only considered medium-resource languages.

Most African languages and indigenous languages from the Americas are nearly absent from AI datasets. The task of compiling data in these languages is extremely arduous, as highlighted by Michael Running Wolf, an American software engineer and AI ethicist who founded Indigenous in AI:

> *"It's a huge, monumental effort. And if you're looking at Montana tribes, like the Northern Cheyenne, no one has that [data]. We're looking at millions of hours of audio, and at best, we have maybe around 100 or so, and some tribes don't have anything."* (as quoted in Mabie, 2023)

2.3 Technical Challenges

These asymmetries entail a range of negative consequences, particularly manifesting as a variety of technical challenges when working with less resourced languages. Firstly, the difficulties arise from the fact that tools designed for English often assume a linguistic structure that involves tokenising or separating sentences into individual words.

However, such tools are not always effective for African languages like isiZulu, which are agglutinative—forming words by stringing together morphemes that each represent a distinct grammatical element, a structure that English-trained AI tools find difficult to process (Ravindran, 2023). This type of issue recurs with many indigenous languages, which are typically polysynthetic; they compact multiple grammatical elements such as subject, object and verb into single-complex words. This linguistic characteristic complicates the use of AI technologies initially developed for languages with simpler morphological structures (Mabie, 2023).

2.4 The Looming Threat of Language Erasure

These problems are not confined to the technical realm; they also have tangible, often severe social consequences. As chatbots, translation devices and voice assistants become increasingly crucial to navigate the web, the rising tide of generative AI could lead to the marginalisation of thousands of indigenous languages that lack sufficient amounts of text to train AI models. David Ifeoluwa Adelani, an expert in NLP for low-resource languages, articulates a vivid scenario highlighting the practical impacts of these technological gaps:

> *"You want to do a task, and you want a machine to do it for you. If you express this in your own language and the technology doesn't understand, you will not be able to do it. A lot of things that simplify lives for people in economically rich countries, you will not be able to do."* (as quoted in Wong, 2024)

Meggan Van Harten, a strategic leader and partner at Design de Plume—an Indigenous-owned, women-led creative agency in Canada—, reflects on how the current shortcomings in applying AI tools to Indigenous languages not only lack accuracy but also directly impact the community, contributing to broader issues. She states,

> *"[this] is the harsh reality that we're living in. The way that we're building out technology, design systems, all of these things, is contributing to a culture of further [cultural] genocide and erasure of languages."* (as quoted in DiBenedetto, 2023)

It is crucial to acknowledge that indigenous languages are experiencing a dramatic decline. For the remainder of the century, it is feared that most of these languages will be lost forever (Dawn et al., 2024). This trend may accelerate with the integration of AI technologies in various aspects of social life. As Michael Running Wolf warns,

> *"Essentially, we're racing against time. Within five to 10 years, we risk losing a significant part of the cultural and linguistic heritage in the United States."* (as quoted in Will, 2024)

2.5 Imbalances in Multilingual Models

At this juncture, it should be noted that many of today's most powerful language models are not monolingual but multilingual—i.e. trained on texts from many different languages simultaneously. Similar to their monolingual counterparts—which are trained through a fill-in-the-blank task—, by training on text from several different languages, multilingual language models can also infer connections between languages. They act as a bridge between high- and low-resource languages, potentially allowing the former to support the development of the latter (Nicholas and Bhatia, 2023: 20).

However, linguistic imbalances also negatively affect these systems. Despite their multilingual capabilities, these models are often still disproportionately trained on English text, thereby inheriting and transferring English-centric values and assumptions into contexts of other languages where they might be inappropriate.

The disparity in data availability means that multilingual language models generally perform better in higher- than in lower-resource languages. Developers sometimes attempt to mitigate these data gaps with machine-translated text, but translation errors can further distort language representation. Additionally, when multilingual language models fail, their complex interlingual connections can make the problems difficult to identify and resolve. Moreover, multilingual language models do not function equally well across all languages because the more languages a model is trained on, the less it can capture the unique characteristics of any specific language. This dilemma is referred to as the "curse of multilinguality" (Nicholas and Bhatia, 2023: 28).

In July 2022, Meta launched its "No Language Left Behind" (NLLB) initiative, training a model on over 200 languages with the goal of "developing high-quality machine translation capabilities for most of the world's languages" (Meta, 2022). Subsequently, in November of the same year, Google introduced its "1,000 Languages Initiative" aimed at "get[ting] much better performance on our low resource languages", according to Zoubin Ghahramani, Vice President of research at Google AI (Vincent, 2022). These ambitious projects, however, have not been without controversy. They have been met with substantial criticism, particularly concerning their adverse effects on local communities involved in AI. This issue is incisively addressed by Asmelash Teka Hadgu, Paul Azunre and Timnit Gebru in the abstract to their article "Combating Harmful Hype in Natural Language Processing":

> *"In recent years, large multinational corporations have made claims of creating 'general purpose' models that can handle many different tasks within natural language processing. Recent works from Meta, for example, give the impression that they have nearly solved machine translation tasks for more than 200 languages, including 55 African languages. In this paper, we outline the harms speakers of non-dominant languages have experienced due to these grandiose and inaccurate claims, ranging from diverting resources from local startups to low-quality datasets and models from these corporations. We urge the African NLP and machine learning communities to push back against these claims and support smaller organisations serving their own communities."* (Hadgu et al., 2023)

The weak presence of low-resource languages in multilingual models not only results in linguistic disparities but also poses cybersecurity risks. A recent study highlights these issues by demonstrating the cross-lingual vulnerability of safety mechanisms in language models. The research shows that by translating unsafe English inputs into low-resource languages, attackers could circumvent the safeguards of GPT-4. This indicates that the limited training on low-resource languages poses a risk to all users of LLMs (Yong et al., 2023).

At the same time, most automatic translation systems, which have utilised deep learning for years, usually exhibit various types of biases, including those related to gender. For example, when the gender-neutral Turkish phrase "*O bir doktor*" is translated into English using Google Translate, it yields both "He is a doctor" and "She is a doctor". Similarly, "*O bir hemşire*" translates to both "He is a nurse" and "She is a nurse". This dual response was adopted by Google in late 2018 (Kuczmarski, 2018), following research that revealed implicit gender assumptions in its translation system—assuming doctors are male and nurses female. Notably, this adjustment was made only for English translations; but in other languages like Spanish or French, translations continue to provide a single, gender-biased response.

Moreover, the shortcomings of many AI-powered translation tools in recognising and understanding regional accents and dialects leads to negative consequences for intercultural communication. For instance, *The Guardian* reports the case of Carlos, an Afro-Indigenous refugee from Brazil, whose mother tongue was Portuguese. During his detention by US Immigration and Customs Enforcement (ICE), the staff attempted to use an AI-powered voice-translation tool to interpret Carlos's statements. However, the system failed to recognise his regional accent. Consequently, Carlos spent six months in detention without being able to communicate effectively, leaving him uncertain about the reasons for his detention and his family's whereabouts (Bhuiyan, 2023).

On the other hand, the widespread use of multilingual LLMs can intensify the risk that the human ability of understanding foreign languages might be relegated to automated intermediaries, thereby losing direct knowledge of these languages and the cultures from which they originate. It is worth noting that in many countries, interest in learning foreign languages is notably declining.

Enrolment in non-English language courses at US universities fell by 29.3 % from 2009 to 2021, according to the Modern Language Association (MLA). Similar trends are observed in Australia, South Korea and New Zealand, where fewer students are studying foreign languages and some universities are closing their related departments (Dans, 2024). This decline may accelerate in the future, further diminishing cognitive and cultural benefits and seriously impacting intercultural understanding.

2.6 Local Initiatives at the Intersection of AI and Languages

In this context, laden with challenges, local initiatives are increasingly pivotal in developing LLMs and other AI tools that cater to the linguistic diversity and specific needs of various countries and communities, encompassing both high-resource and low-resource languages. Below, some examples from different countries are provided.[5]

One such project is *Lengua Española e Inteligencia Artificial*, otherwise known as LEIA,[6] spearheaded by the Royal Spanish Academy, which aims to defend, promote and ensure the correct usage of the Spanish language in the digital world, particularly in the age of AI. The primary objectives include overseeing the proper use of Spanish in machines and utilising AI to foster its correct use among humans. Continuing the focus on Spain's efforts at the intersection of AI and languages, the country is poised to develop an open-source LLM that will be trained not only in Spanish/Castilian but also in Basque, Catalan, Galician and Valencian (Gallardo, 2024).

BigScience's Large Open-science Open-access Multilingual Language Model (BLOOM) project[7] serves as another example of how LLMs can be developed more transparently and with extensive public support. Funded in part by the French government among other stakeholders, BLOOM has maintained an open policy, allowing continuous inquiry and analysis by NLP practitioners

[5] It should be noted that this list is not exhaustive, and new cases are emerging on a monthly basis.

[6] https://www.rae.es/leia-lengua-espanola-e-inteligencia-artificial.

[7] https://bigscience.huggingface.co/blog/bloom.

worldwide. This LLM utilises a 1.6 TB multilingual dataset that has been meticulously documented, making it readily accessible for NLP research and development (Le Scao et al., 2022).

Common Corpus,[8] also supported by the French Ministry of Culture and coordinated by the French start-up Pleias, involves a consortium including Allen Institute for AI and EleutherAI. It has produced what is considered the largest open AI training dataset in the public domain, aiming to provide researchers and start-ups with high-quality curated data.

The First Languages AI Reality (FLAIR) initiative,[9] led by Michael Running Wolf at the Mila Institute, focuses on using AI to preserve endangered indigenous languages. FLAIR emphasises building trust within communities to ensure that sacred and culturally sensitive materials are respected and used appropriately in AI applications.

Silo AI, based in Helsinki, has developed the Poro model[10] in collaboration with the University of Turku. Silo AI integrates Danish, Finnish, Icelandic, Norwegian and Swedish into its open-source models, focusing on the nuances of these languages and ensuring that LLMs reflect the values and cultures of the Nordic people.

In Africa, grassroots organisations like Masakhane[11] and Lelapa AI[12] are at the forefront of integrating local languages into AI. Masakhane (which can be translated to "We build together" in isiZulu) advocates for NLP research in African languages by Africans, challenging the notion that the lack of data is merely a technical problem, instead framing it as a societal challenge. Lelapa AI, inspired by the philosophy of Ubuntu, focuses on leveraging African expertise, positioning itself as a nurturing hub for African AI talent to address both local and global issues.

[8] https://huggingface.co/collections/PleIAs/common-corpus-65d46e3ea3980fdcd66a5613.
[9] https://mila.quebec/en/first-languages-ai-reality.
[10] https://huggingface.co/LumiOpen/Poro-34B.
[11] https://www.masakhane.io.
[12] https://lelapa.ai.

3. AI in the Creative Sectors: Applications and Challenges

3.1 Opportunities Across the Cultural Value Chain

Culture is one of the sectors most impacted by AI technologies. Prior to the rise of generative AI, the technologies were already extensively used to recommend online content, tailoring cultural consumption to individual preferences and enhancing user engagement. Now, with the advent of LLMs, diffusion models and other generative systems, it has become clear that culture serves not only as an input for machines—which process texts, melodies, videos and other forms of media—but also as a significant output.

The impact of AI systems is pervasive across all creative sectors. In literature and publishing, writers benefit from a new generation of AI assistants and chatbots, which aid in overcoming a writer's block, refining passages and providing other writing aids. Publishers utilise AI tools to analyse lengthy texts, automatically correct errors, translate sections, generate summaries and categorise works, thus greatly improving efficiency.

In the music industry, applications such as AIVA, Udio and Suno allow users to compose high-quality songs without extensive knowledge of arrangements and orchestration. Additionally, Adobe's new AI tools are revolutionising music editing, likened to the 'Photoshop' for music production (Weatherbed, 2024).

The visual arts have also undergone a transformation, with a proliferation of AI models that enable the creation of high-realism images in seconds. These services also allow for rapid editing of image components, dramatically reducing the time needed for detailed artwork.

The film industry is not far behind, with emerging applications that generate high-quality video content from simple prompts. Video-editing tools are becoming more and more sophisticated, such as Viggle, which allows users to replace any individual in a video with just a few clicks.

In the world of video games, AI technologies are used not only in game development but also enhance player interactions and game dynamics, making the gaming experience more immersive. These tools enable characters and

scenarios to adapt in real time to player actions. Additionally, they improve the intelligence and realism of non-player characters (NPCs) allowing for more complex interactions. Many of these advancements are evident in games such as *The Sims*, *No Man's Sky*, *Minecraft* and *Fortnite*.

In the GLAM (Galleries, Libraries, Archives and Museums) sector, AI technologies offer substantial benefits: they can be used to digitise and catalogue vast collections, which enables interactive and immersive virtual experiences that make cultural heritage more accessible (see for example Kobberod, 2024). Additionally, the use of generative AI can help digitally reconstruct lost or damaged artifacts.

The impact extends even into sectors traditionally less associated with digital technology, such as the Performing Arts. For instance, the creation of virtual avatars in K-pop has been revolutionary. AI-generated characters perform just like human bands but without physical limitations, participating in live chats, videos and fan interactions (Kwon and Watson, 2023).

As can be seen, AI technologies are profoundly transforming the cultural landscape, leaving no field untouched. Broadly speaking, this new era heralds several significant shifts: it can enhance the creative potential of artists, introducing the concept of the "augmented artist" who leverages technology to push traditional boundaries. It also indicates the emergence of new roles at the intersection of art and technology, such as artist-programmers, who blend programming skills with an artistic vision. There is also a substantial increase in productivity in the cultural and creative industries. And last but not least, the public finds itself with access to a vastly greater array of cultural works than ever before.

3.2 Opacity, Lack of Diverse Datasets and Low-Quality Content

However, it is also necessary to pay attention to the new challenges that arise from these technologies. Indeed, their widespread introduction into the creative ecosystem could result in increasing homogenisation and a marked

impoverishment of cultural diversity. At the same time, the growing difficulties faced by artists and creative entrepreneurs in the age of AI could irreversibly weaken an essential source of cultural vitality.

Firstly, there is a significant concern regarding the lack of transparency in content recommendation algorithms, which are commonly found in various digital platforms and services that curate and suggest content to users, such as streaming services, social media platforms and e-commerce sites. These systems, often described as "black boxes," analyse user behaviour and preferences to recommend movies, music, posts and more, aiming to enhance user engagement. The internal mechanisms of these algorithms are typically not disclosed, which raises concerns about the potential for bias. This opacity often impacts the discoverability of local content and does not necessarily promote a broader exchange among cultures. Moreover, the fact that training data for generative AI predominantly comes from American, European, or Chinese sources can also lead to a lack of diversity in future creations (Usbek & Rica, 2023).

Additionally, at a time when the barriers to using generative AI have been lowered, there has also been a proliferation of fake, low-quality artistic content, particularly "fake books"—i.e. AI-generated texts that mimic the style and format of genuine publications but lack originality and depth, often misleading consumers. This problem has become so severe that in September 2023, Amazon was compelled to introduce a new policy for Kindle authors, limiting them to publishing a maximum of three books per day on the platform (Creamer, 2023).

3.3 Intellectual Property

Another highly controversial issue in the field of generative AI relates to intellectual property. For starters, it remains unclear for now as to who holds the rights to a work created with automated tools like Midjourney or Dall-E, as many parties are involved: the end user who devised the prompt, the company that built the model and thousands of people whose original works were used to train it. Most courts thus far have ruled against attributing authorship to AI itself, as authorship and inventions should apply to a human person (see for example regarding the US: Brittain, 2023; regarding the UK: Jackson et al., 2024).

In any case, the fact that is truly problematic in most current generative AI applications is that there is not enough transparency in terms of input data and even less compensation for the rights holders—artists, agencies, production companies, among others. Remarkably, AI companies do not even conceal the fact. In its presentation before the UK House of Lords Communications and Digital Select Committee inquiry in December 2023, OpenAI admitted that:

> *"Because copyright today covers virtually every sort of human expression—including blog posts, photographs, forum posts, scraps of software code, and government documents—it would be impossible to train today's leading AI models without using copyrighted materials."* (OpenAI, 2023: 4)

A survey from January 2024 by the UK-based non-profit organisation Design and Artists Copyright Society (DACS) found that 22 % of 1,000 surveyed artists and agents had already discovered that their own work had been used to train AI (Tapper, 2024). In many cases, AI companies have resorted to datasets such as Books3, which contain many pirated books, leading to several authors taking legal actions (Reisner, 2023).

Tech firms often invoke fair use to justify their practices; however, this rationale has not garnered much support. In November 2023, Ed Newton-Rex made headlines in the generative AI industry when he announced his resignation from his position as Head of Audio at Stability AI, due to the company's stance that it was acceptable to use copyrighted work without permission to train its products. His argument was compelling:

> *"Today's generative AI models can clearly be used to create works that compete with the copyrighted works they are trained on. So I don't see how using copyrighted works to train generative AI models of this nature can be considered fair use."* (Newton-Rex, 2023)

Indeed, if generative AI techniques lack transparency, they can undermine the sustainability of creators' work, facilitate the unauthorised replication of contemporary artists' styles and even lead to the cultural appropriation of a people's intangible heritage.

After resigning from Stability AI, Newton-Rex founded the Fairly Trained project,[13] which offers certificates for generative AI companies that obtain consent for their training data. Concurrently, tools have proliferated to assist creators in locating unauthorised uses of their works by AI systems, such as CoverNet.[14] Furthermore, developments in data poisoning techniques have emerged, such as Nightshade,[15] which allows artists to subtly alter their works in ways imperceptible to the human eye but that can severely disrupt computer vision, leading to distorted outputs. Finally, there are many who advocate for the use of "algorithmic disgorgement" to AI companies—a technical term for the enforced deletion of algorithms developed using illegally collected data.

In this context, many in the creative sectors are advocating for new compensation schemes for artists, based on licensing systems. Mary Rasenberger, chief executive of the Authors Guild—the oldest and largest professional organisation in the United States for published writers—explains:

> *"We have to be proactive because generative AI is here to stay. They need high-quality books. Our position is that there's nothing wrong with the tech, but it has to be legal and licensed"* (as quoted in Cho, 2024).

However, it is worth questioning whether this model would function effectively beyond the borders of the Global North, as creators from developing countries often lack access to international agencies and are not typically members of collective rights-management societies.

3.4 Job Loss in the Creative Sectors

The impact of AI, particularly generative AI, is already being felt in terms of economic activity. While the cultural and creative industries are leveraging many of the advantages offered by AI technologies, the use of these tools threatens to push many players out of the market. For example, Tyler Perry, a prominent

[13] https://www.fairlytrained.org.

[14] https://www.covernet.ai.

[15] https://nightshade.cs.uchicago.edu.

American actor, filmmaker and entrepreneur, had been working on a substantial expansion plan of about 800 million US dollars for his Atlanta-based studio. However, the launch of OpenAI's Sora, known for its advanced cinematic video outputs, has been a significant factor in Perry's decision to temporarily halt the project (Kilkenny, 2024).

The negative impact on employment is already widely noticeable. In July 2023, a study using data from the freelancer platform Upwork showed that copywriters and graphic designers experienced substantial declines in both job opportunities and earnings, indicating that generative AI not only replaces their jobs but also devalues their work (Hui et al., 2023).[16] Along the same lines, a survey by the UK Society of Authors from January 2024 indicated a substantial impact of generative AI on creative careers. About a quarter of illustrators and over a third of translators reported job losses, with many also experiencing a decline in income value. Additionally, a majority of writers voiced fears of future negative effects on earnings, particularly translators and illustrators, with over three quarters expressing a concern.

On the other hand, a study conducted in late 2023 by GEMA and SACEM—collective management organisations based in Germany and France, respectively, representing authors, composers, music publishers, among others—highlighted similar trends in the music industry. Their survey, carried out among their members, projected that by 2028, generative AI could endanger 27 % of music creators' revenues, potentially causing financial losses of approximately 950 million euros in that year alone, with a cumulative impact of around 2.7 billion euros over five years (Goldmedia, 2024: 83).

In 2023, Hollywood writers went on a 5-month-long strike, driven by fears that studios would use generative AI to replace them, degrade writing jobs, exacerbate financial insecurity and alter the career ladder, especially after recent disruptions from streaming platforms. The writers' concerns united them during the gruelling strike that actually lasted 148 days and ultimately led to a historic

[16] The study utilised a difference-in-differences design to assess the impact of generative AI on freelancer employment outcomes. This approach analysed changes in employment and earnings before and after the introduction of AI technologies like ChatGPT, effectively controlling for unobserved, time-invariant differences among individuals.

victory in their new contracts with studios. The agreement gives writers agency over AI use, ensures that they receive full credit and compensation, and it prohibits studios from forcing writers to employ AI, setting an important precedent for other unions (Kinder, 2024).

4. Navigating the Information Environment: Synthetic Content, Deepfakes and Beyond

From the standpoint of access to information, communication and the exchange of ideas, it is evident that AI technologies can offer significant benefits. For instance, AI systems enable media companies the sifting-through of massive databases, the enhancing of speed and accuracy of research and reporting, and the full leveraging of the capabilities of LLMs in NLP tasks. For social media platforms, AI is a powerful tool for content moderation, as well as for fostering greater engagement and delivering relevant content to each user.

In the political arena, the use of AI systems could transform democratic processes by educating citizens, aiding deliberations, summarising views and identifying common grounds. LLMs could help politicians understand their constituents' needs and assist the public in articulating political positions more clearly. Additionally, these models could also serve as educators on policy issues, facilitating a deeper individual understanding and potentially supporting large-scale democracy through online discussions (Schneier et al., 2023).

Indeed, a 2023 study demonstrates that LLMs can significantly enhance the quality of these exchanges by making them more productive, reducing incivility and fostering willingness to engage in challenging conversations across social divides. In a detailed experiment, researchers used GPT-3 to provide real-time suggestions for rephrasing messages in conflict-mediation style on a custom-built chat platform, where participants discussed divisive topics like gun control. This intervention not only improved the perceived quality of conversations but also decreased toxicity by 15 %, without changing the fundamental content of the discussions or shifting participants' policy stances (Bail et al., 2023).

4.1 The Proliferation of Synthetic Content

However, as is the case in other fields, the application of AI technologies also presents challenges in the realms of information and communication. To begin with, the surge in synthetic content is saturating the web with low-quality materials, a process many liken to the pollution of oceans with plastic trash or the atmosphere with carbon dioxide (Anderson, 2023). This not only risks leading to a potential model collapse, as mentioned in Chapter 1—since future

AI systems will be trained on these data of limited added value—but also to a degradation of the information environment.

It is important to note that synthetic content may not only be redundant or unoriginal but may also be laden with various stereotypes and errors. A study in April 2024, conducted by NewsGuard, an organisation that tracks misinformation, identified hundreds of AI-generated news websites operating with little to no human oversight (Sadeghi et al., 2024). These websites, bearing generic yet credible names like "iBusiness Day," "Ireland Top News" and "Daily Time Update," create the illusion of legitimacy while depending on automated systems to produce a vast array of content across various topics such as politics, technology, entertainment and travel, often featuring inaccuracies, including false claims about public figures and outdated events portrayed as current.

4.2 Misinformation, Biases and Manipulation

The issue of misinformation is not limited to obscure sites but also affects automated news services from leading companies that source content from this type of portal. Notable errors have been transmitted by Microsoft's MSN.com; for instance, false claims have been circulated about President Joe Biden falling asleep during a moment of silence for wildfire victims, conspiracy theories about the Democratic Party orchestrating a surge in Covid-19 cases and an offensive obituary of an NBA player (O'Sullivan and Gordon, 2023).

The risk of delivering false information is also present in search engines and AI-powered chatbots, and can have profound and detrimental effects on democratic processes. A study on Microsoft's Bing Chat, now known as Microsoft Copilot, highlighted significant issues in delivering accurate election-related information for the Bavarian, Hessian and Swiss elections in October 2023. From August to October 2023, researchers discovered that a third of the AI's responses contained factual errors, including wrong election dates, outdated candidate details or invented controversies. The system's safeguards were inconsistently applied, leading to evasive responses in 40 % of the cases and failure to adequately address direct questions about candidates. These persistent inaccuracies over time threaten to undermine the credibility of both political figures and media outlets (AI Forensics, 2023).

The challenges extend beyond mere factual errors to include biased responses. In early 2023, a software engineer in Bengaluru launched GitaGPT, a chatbot inspired by the *Bhagavad Gita* and designed to emulate the tone of the Hindu god Krishna. This bot and similar versions raised concerns about promoting harmful ideologies, as some lacked filters against casteism and misogyny, and even suggested that certain violent actions could be justified as one's duty. Similarly, chatbots based on the Quran stirred controversy by delivering extreme advice, leading to one being paused and another being shut down after a community backlash (Nooreyezdan, 2023).

While some chatbots theoretically have indeed the potential to foster more civil and respectful communication, in practice, they can become powerful engines of misinformation, propaganda and manipulation. Several studies have shown that LLMs have reached, or even surpassed, human levels of persuasion, regardless of the factual accuracy of their outputs (Rescala et al., 2024; Spitale et al., 2023). This issue becomes even more concerning in a context where AI systems are increasingly adept at discerning user preferences, as noted by Sam Altman, CEO of OpenAI:

> *"A thing that I'm more concerned about is what happens if an AI reads everything you've ever written online … and then right at the exact moment, sends you one message customised for you that really changes the way you think about the world."* (as quoted in Shah, 2023).

During election times, the impact of AI-enhanced misinformation—both in generating fake content and in targeting the right audience—can be profound, particularly in years like 2024, with elections taking place in over 50 countries. In this context, some analysts predict a "microtargeting tsunami" (Heath, 2023), accompanied by a proliferation of strategies aimed at manufacturing the perception of consensus around political issues, undermining government responsiveness, swaying public opinion, exacerbating divisions, demobilising or deceiving voters and eroding trust in electoral processes (Wirtschafter, 2024).

4.3 AI Companies and Media Outlets

The proliferation of synthetic content has underscored the value of high-quality, human-generated content, as exemplified by articles from reputable press outlets. In this scenario, companies like OpenAI have attempted to secure significant deals to utilise journalistic content for training their AI models. For instance, OpenAI has negotiated a multi-million-euro agreement with German publisher Axel Springer SE for access to content from its news platforms, including *Politico, Business Insider, BILD* and *WELT*, and has also reached an agreement with AP to use its news archive (Nguyen, 2024).

However, even as media companies are negotiating licensing deals with AI firms, they are simultaneously establishing "digital blockades." Recent data shows that over 88 % of top-ranked US news outlets now prevent web crawlers employed by AI companies from accessing their content for use in training chatbots and other AI projects (Knibbs, 2024). Some organisations, like *The New York Times*, have gone even further by taking legal action against OpenAI to prevent their content from being used to train models like ChatGPT.

It is noteworthy that right-wing media outlets are significantly less proactive in implementing bot-blocking measures compared to their liberal counterparts, which could potentially influence the outputs of AI systems, reflecting a skewed data environment.

4.4 A New Era of Searching

When it comes to searches, a paradigm shift is underway. Instead of displaying the traditional list of web-sourced search results, major search engines are migrating towards a new model based on generative AI. Google, which dominates more than 91 % of the search market, is implementing a profound change to its interface called the "Search Generative Experience" (SGE). This service employs an AI that amalgamates facts and text snippets from various sources, often replicating them verbatim, and presents this aggregated content as its main output.

As generative search prioritises the direct provision of information—reducing the need for users to click through to external sites—the new model could drastically affect many publishers who rely heavily on Google for traffic, potentially leading to decreased visibility and revenue (Piltch, 2023). This transition could inadvertently solidify the positions of large media companies with established business models, while making smaller and independent sites economically unfeasible.

The changes will also profoundly affect searching as a human experience. In a "traditional" web search, users find results from external pages—most often authored by human beings—which they can then choose to visit based on their interests. Now, with generative search and generally with any chatbot interface, users no longer confront human "otherness" but rather the machine's "alienness," marked by its specific opacity. This type of change represents a qualitatively new type of interaction, transitioning from dealing with a "them" to an "it" (Garcia, 2024). In this way, there is a risk of replacing encounters with real diversity—the content created by other human beings, with their vast plurality of voices and perspectives—by a much more limited experience, in which the user simply accesses a summary of multiple sources which the machine has selected and, rather opaquely, "digested."

4.5 Social Media, Content Moderation and Labelling

For several years now, social media platforms have been leveraging AI for content moderation, particularly to identify and manage posts that propagate disinformation and hate speech. Despite these efforts, the algorithmic tools still struggle to perform their duties effectively. Indeed, cultural context and language present substantial obstacles for content moderation systems. Similar to many of the models studied in Chapter 2, these tools are primarily trained on English data and often underperform when they have to deal with other languages. For example, while Twitter can automatically block offensive tweets in English, it lacks the same capability for African languages (Ravindran, 2023).

Due to the context-specific nature of language and cultural nuances, major social media platforms incorporate a human-in-the-loop approach, involving human experts to oversee technical moderation processes. It is particularly apparent that especially the local level needs human moderation to be effective. Moderation teams must understand local languages, slang and subtle cues that may be regionally specific.

However, the content moderation teams of American social media companies speak predominantly English. This is problematic, as revealed by *The Wall Street Journal*, which found that in 2020, 87 % of content moderation efforts were focused on posts from the US—even though it was also an election year in numerous other countries like Egypt, Poland, Sri Lanka and Tanzania. This discrepancy highlights a severe imbalance, considering that 90 % of Facebook users reside outside the US (Chakravorti, 2023).

Despite these concerns, social media companies are actually cutting back on their content moderation teams. In 2023, Meta reportedly disbanded its Responsible AI team as it shifts its focus on generative AI (Huand and Varnham O'Regan, 2023). Concurrently, under Elon Musk's leadership, Twitter (now X) has drastically reduced its moderation staff, eased restrictions and reinstated previously suspended accounts.

Various platforms are implementing technical measures to identify synthetic content, centring these efforts around content labelling and technological safeguards. YouTube now requires users to indicate something as being "synthetic media" when uploading videos. Meta has expanded its labelling to include a broader range of content marked as "Made with AI" when AI involvement is detected or specified by users. TikTok is developing tools to assist creators in labelling their AI-generated content and is experimenting with automated labelling techniques. OpenAI is also promoting transparency by using digital credentials to verify content origins. Furthermore, OpenAI has set boundaries in its Dall-E image generator, rejecting requests for images of real individuals, like political figures, and preventing the creation of chatbots that impersonate real people or institutions.

However, these systems are not without their limitations. For example, the effectiveness of many image labelling systems often depends on images

previously having been marked with AI-generated watermarks in their metadata. Yet, there are tools available that can bypass these metadata markers (Lima-Strong and Oremus, 2024). Additionally, systems designed to automatically detect AI-generated content are far from perfect when it comes to videos and images and are even less reliable with texts.

4.6 Online Gender Violence

Concerning technology-facilitated gender-based violence (TFGBV), estimates from 2020 indicate that 58 % of young women worldwide have experienced some form of gender-based violence on social media platforms (Chowdhury, 2024: 11). Both open and closed AI models can create cyber-harassment templates, synthesising false backgrounds for individuals and altering images to depict people in various non-consenting scenarios. These capabilities highlight the potential for generative AI to amplify common TFGBV harms, including impersonation, hacking, stalking and cyber-harassment (Chowdhury, 2024).

Taylor Swift has become a high-profile victim of deepfake pornography, a growing issue as generative AI advances. In just the first nine months of 2023, 113,000 deepfake videos were uploaded to popular porn sites, compared to 73,000 in 2022, with a study showing that 96 % of internet deepfakes are pornographic in nature (Hoover, 2024).

4.7 The Role of Big Tech Companies

In practice, many come to recognise that current moderation methods—whether algorithmic or hybrid, involving both humans and machine—have not advanced due to the lack of cooperation from tech companies. According to Hany Farid, a professor at the University of California, Berkeley, School of Information,

> *"Content moderation has not kept up with the threats because it is not in the financial interest of the tech companies. This is all about greed. Let's stop pretending this is about anything other than money."* (as quoted in Ryan-Mosley, 2023)

In September 2022, Nobel laureates Maria Ressa and Dmitry Muratov launched a 10-point action plan to address the threats posed to democracy by big tech's business model. Their plan advocates for the ending of the surveillance-for-profit model, establishing robust data protection laws and rebuilding independent journalism to counter the influence of major technology firms:

> *"We urge rights-respecting democracies to wake up to the existential threat of information ecosystems being distorted by a big tech business model fixated on harvesting people's data and attention, even as it undermines serious journalism and polarises debate in society and political life."* (Ressa, Muratov et al., 2022)

4.8 The Liar's Dividend

In any case, in a world increasingly filled with synthetic, false or dubious content, "you can't trust anything you see or hear," as expressed by Eric Schmidt, former CEO of Google (Chakravorti, 2023). There is even the possibility that one might not know whether they are interacting with a person or a bot. This environment fosters a certain phenomenon, the so-called "liar's dividend," where any piece of evidence can be dismissed as a potential deepfake, further undermining trust (Bontcheva, 2024). In such circumstances, there is a risk of "false positives," where legitimate content is mistakenly removed, raising serious questions about freedom of expression. In fact, the accusation of "fake news" has sometimes been used to curtail the activities of independent media; for instance, in 2022, out of 363 journalists incarcerated globally, 39 were detained under accusations of "fake news" or for violating disinformation policies (CNTI, 2024: 2).

4.9 Censorship in the Age of AI

In many countries, AI systems have long been used to enhance government censorship and surveillance capabilities (Feldstein, 2019), often with the complicity of major tech companies. However, generative AI brings these issues to a whole new level. Censorship can now be applied at the input stage of an AI system, requiring developers and users to monitor the types of data used to train

machines, ultimately affecting outputs. For instance, in China, AI tools typically source their information from within the so-called Great Firewall, such as Baidu's Baike encyclopaedia, rather than the uncensored Chinese version of Wikipedia, resulting in outputs that mirror the country's heavily censored information landscape. Conversely, censorship can also be directly applied to outputs. For example, when Baidu launched its text-to-image generator in 2022, users quickly noticed gaps and manipulations, indicating that the content had been extensively censored (Cook, 2023).

It is important to note that censorship within a country not only impacts domestic conditions but can also have global repercussions. In 2022, a report by the Canadian research group Citizen Lab revealed that Microsoft's Bing had engaged in the persistent censorship of politically sensitive Chinese names. This censorship extended across multiple Chinese political topics, was conducted in at least two languages—English and Chinese—and affected users in various regions, including China, the US and Canada (Knockel and Ruan, 2022).

5. Tech Titans: Ambitions and Imaginaries at the Helm

5.1 Main Actors

To fully grasp the prevailing trends in the AI landscape, it is critical to identify the major players driving these advancements and understand their mindset.

OpenAI, established in 2015, initially started as a not-for-profit organisation but has evolved significantly, especially after creating for-profit subsidiaries in 2019. This restructuring was aimed at adapting to an inflow of 1 billion US dollars from Microsoft and other tech companies to accelerate its global impact. By January 2023, Microsoft deepened its commitment with an investment of an additional 10 billion US dollars, boosting OpenAI's valuation to 80 billion US dollar (Metz and Mickle, 2024). Under the leadership of CEO Sam Altman, OpenAI has developed landmark applications like ChatGPT and Dall-E.

Anthropic, established in 2021 and co-founded by siblings Dario and Daniela Amodei, along with other former OpenAI employees, has rapidly emerged as a key player in the AI sector. Supported by substantial investments, including a significant 4 billion US dollars from Amazon, the company unveiled the third iteration of its Claude model in March 2024, positioning itself as a direct competitor to OpenAI's ChatGPT with comparable capabilities in processing speed and accuracy (Tyler Millward, 2024).

DeepMind, established in 2010 in London by Demis Hassabis, Shane Legg and Mustafa Suleyman, rapidly advanced the field of AI and was acquired by Google in 2014 for around 500 million US dollars (Perrigo, 2023a). Renowned for creating AlphaGo, the first AI to defeat a professional Go player, DeepMind has significantly contributed to Google's AI projects, including the development of the Gemini chatbot, formerly known as Bard.

Inflection AI, co-founded in early 2022 by Mustafa Suleyman and LinkedIn's Reid Hoffman, positioned itself as a key innovator in the AI sector with its focus on creating more personalised AI through its product, Pi. The company garnered significant industry attention by securing 1.3 billion US dollars in funding in 2023 from major backers including Microsoft and Nvidia. However, the start-up's journey took a dramatic turn when Microsoft, less than a year later, integrated Inflection's team and developments into its own AI division (Hu and Varghese, 2024).

Hugging Face, founded in 2016 by French entrepreneurs Clément Delangue, Julien Chaumond and Thomas Wolf in New York City, has evolved from a chatbot to a major player in open-source AI. With over a million models, datasets and apps, Hugging Face supports a vibrant community akin to a "GitHub for AI," offering a platform for the innovation and collaboration of over 10,000 companies. The platform has attracted heavyweights like Google and Microsoft, boosting its valuation to 4.5 billion US dollars in 2023 (Leswing, 2023).

Stability AI, a London-based company established by Emad Mostaque in 2019, rapidly received recognition for its pioneering open-source AI technologies, notably its popular text-to-image model Stable Diffusion, released in 2022. The company initially relied on self-funding but later attracted substantial venture capital investments, totalling over 100 million US dollar. However, by early 2024, Stability AI encountered operational hurdles that impacted its financial position, valuation and investor relationships (Mathews and Garfinkle, 2024).

Midjourney, founded in 2021 by David Holz in San Francisco, has quickly become a leader in generative AI for image creation. This bootstrapped start-up —meaning it has been funded by its own operating revenues—has achieved 200 million US dollars in annual recurring revenue by 2023 without venture capital funding (Clark, 2023).

Mistral AI, founded in 2023 in Paris by Arthur Mensch (a former DeepMind employee) along with Guillaume Lample and Timothée Lacroix (both formerly at Meta), is dedicated to developing open-source LLMs. The company quickly secured significant financial support, raising 428 million US dollars by October 2023, which pushed its valuation to over 2 billion US dollars with promising future prospects. In February 2024, Microsoft announced an investment in the company (Dillet, 2024).

Aleph Alpha, founded in 2019 in Heidelberg by Jonas Andrulis (a former Apple employee) and Samuel Weinbach, also specialises in LLMs. The company secured over 500 million US dollars in a funding round in 2023 led by Innovation Park Artificial Intelligence, Bosch Ventures and other prominent investors (Browne, 2023).

There are countless AI start-ups outside the US and Europe, most notably in China. **Baichuan Intelligence**, founded by former Sogou executives Wang Xiaochuan and Ru Liyun in Beijing, has developed its own open-source models and raised 300 million US dollars from giants like Alibaba and Tencent (Dotson, 2023). **Lightyear AI**, initiated by Wang Huiwen, has attracted significant attention and investment from major players, including Tencent, before its acquisition by Meituan (Jiang, 2023). **Zhipu AI**, a spinoff from Tsinghua University founded in 2020, has rapidly gained prominence with its bilingual ChatGLM model, securing 350 million US dollar from top investors including Alibaba (Peng, 2023). Finally, **01.AI**, launched by former Google China president Kai-Fu Lee, is advancing open-source AI by releasing models capable of handling both text and images, with substantial backing from industry heavyweights such as Alibaba.

5.2 Lack of Diversity

Behind the diverse array of international projects developing AI services and products, several trends emerge. First, the leaders of these projects often form a relatively insular network, with frequent transitions between companies indicating a somewhat closed loop of professional mobility.

Second, the leadership in AI enterprises and projects across various regions is predominantly composed of young males. This trend extends to other team members as well, perpetuating the gender imbalances typically observed in AI research. Indeed, research conducted in 2018 by *WIRED* and Montreal startup Element AI assessed the global diversity among leading machine learning researchers and found that only 12 % were women (Simonite, 2018). Furthermore, in 2019, the World Economic Forum reported that women made up only about 26 % of the workforce in data and AI roles (Crotti et al., 2019: 40). The lack of diversity is also evident when it comes to race—in the US, Black professionals comprise only 7.4 % of the tech workforce (Jackson, 2022).

Additionally, regarding professional backgrounds, many tech leaders possess degrees in fields like computer science or engineering, which contrasts sharply with the underrepresentation of those with backgrounds in humanities, social sciences or arts.

5.3 Open-Source vs. Proprietary Models

Another visible trend is the growing divergence between open-source and proprietary models. The rivalry between proponents of these two approaches in the industry sometimes reaches notable intensities. As discussed in Chapter 1, in the debates surrounding AI safety, proponents of closed/proprietary models like ChatGPT argue that these models reduce the risk of rogue actors misusing the technology or AI systems causing uncontrollable harm. Figures like Sam Altman advocate for a licensing-based model, in which AI service providers would need to register and obtain a government permit, accessible only to a few certified companies. This stance has ignited fervent responses. Andrew Ng, co-founder of Google Brain—subsequently integrated into Google DeepMind—, former chief scientist at Baidu and professor at Stanford University, warns against those who stoke fear about AI threats "to argue for legislation that would be very damaging to the open-source community" (Davidson, 2023). Similarly, Yann LeCun, Chief AI Scientist at Meta, also expresses a strong opposition:

> *"Altman, Hassabis, and Amodei are the ones doing massive corporate lobbying at the moment. They are the ones who are attempting to perform a regulatory capture of the AI industry [...]. Openness is the only way to make AI platforms reflect the entirety of human knowledge and culture."* (LeCun, 2023)

However, some suggest that this defence of open-source models may align with specific corporate interests. For instance, Meta puts one of its biggest bets on the Metaverse, where the company stands to benefit from having thousands of developers create artificial environments and various types of content for this new landscape. To facilitate this, Meta needs to distribute powerful open-source models like Llama (Kulveit, 2023).

5.4 The Influence of Major Tech Companies

Another distinctive phenomenon in the current AI industry landscape is the rarity of start-ups remaining bootstrapped; indeed, the vast majority rely on significant injections of capital that need periodic renewal. This trend has led

many analysts to liken the explosive growth of AI applications to a bubble—i.e. a market situation where the prices of assets inflate rapidly, driven more by speculative expectations than by underlying fundamentals—, from which only a few projects are likely to emerge successfully (Doctorow, 2023a). Moreover, the main source of these investments typically comes from large tech companies, to such an extent that when a start-up is either very successful or, paradoxically, when it faces failure, it often ends up being absorbed by the digital behemoth that has supported it—as occurred with Inflection and Microsoft.

In this context, both the overtly and subtly major forces shaping the AI scene are essentially the big tech firms, such as Google/Alphabet, Microsoft, Amazon, Facebook/Meta, Apple, X/Twitter, Baidu, Alibaba, Tencent, ByteDance, Huawei and Nvidia. These companies not only invest in external projects—some of which are eventually integrated into their core business—but also heavily embed AI technologies into their products and refine their own models. Notable examples include Google with Gemini, Microsoft with Copilot, Meta's open-source Llama, X's Grok bot, Baidu's Ernie and other systems that dozens of millions of users globally interact with on a daily basis.

5.5 Monopolisation, Massive Job Loss and Inequality

In this scenario, one of the emerging risks is the reinforcement of monopolistic or oligopolistic conditions by tech titans who control the majority of AI talent, infrastructure and tools. A recent *Brookings* analysis makes it clear that the inherent dynamics of foundation models typically result in a market concentration. It highlights that the high fixed costs of training these models and the low marginal costs of deployment foster significant economies of scale. The report points out that it is more cost-effective for an AI company to produce various models for different uses than for multiple companies to undertake these efforts separately. The marketplace for foundation models also enjoys first-mover advantages—i.e. the benefits gained by the initial significant occupant of a market segment—, although these advantages necessitate substantial ongoing investments. Other barriers such as scarce resources like talent, data,

computational power and intellectual property rights also contribute to a tendency toward natural monopolies in this segment (Vipra and Korinek, 2023).

Chapter 3 mentioned statistics regarding job losses in the creative sector. However, it should be noted that employment disruptions resulting from the integration of AI technologies will span across all economic sectors. According to the International Monetary Fund (IMF), 60 % of jobs in advanced economies like the US and UK are susceptible to AI, with potentially half of these jobs facing negative impacts. In emerging economies such as Brazil, Colombia, India or South Africa, the exposure to AI-related job disruptions is 40 % and 26 % in low-income countries, with a worldwide average of just beneath 40 %. The IMF posits that younger workers, being adaptable and well-versed in new technologies, might seize new opportunities more readily. Conversely, older workers may struggle with reemployment and mobility (Cazzaniga et al., 2024).

A report by the Institute for Public Policy Research (IPPR) suggests that an AI-induced job crisis could erase nearly 8 million positions in the UK. The initial wave of AI adoption is already compromising jobs as more organisations incorporate this technology. A subsequent wave could result in further job automation as AI systems evolve to manage more complex tasks. The study scrutinised 22,000 tasks across various job categories and discovered that 11 % of these tasks—currently performed by workers—are in jeopardy. This risk could escalate to 59 % in the next phase of AI development (Jung and Srinivasa Desikan, 2024).

With millions of individuals being displaced from the labour market and a few tech companies achieving extraordinary valuations, it is unsurprising that inequality may escalate to unprecedented levels. In this context, an insightful incident occurred during a meeting of DeepMind's ethics board in August 2015—before the establishment of Inflection AI, yet after Google's acquisition of DeepMind—in which Mustafa Suleyman introduced a discussion titled "The Pitchforkers Are Coming" (Metz et al., 2023b). He voiced concerns that, as technology could replace numerous jobs in the future, the public might perceive Google as appropriating their means of livelihood.

5.6 Power Concentration

The implications of these trends extend far beyond the realms of economy and employment, which are already significant reasons for concern. As highlighted by Amba Kak, Executive Director, Sarah Myers West, Managing Director of the AI Now Institute—a New York-based policy research organisation specialising in AI—, and Meredith Whittaker, President of Signal and Chief Advisor of the Institute, these changes impact a broader spectrum of societal dimensions:

> *"Concentrated power isn't just a problem for markets. Relying on a few unaccountable corporate actors for core infrastructure is a problem for democracy, culture, and individual and collective agency. Without significant intervention, the AI market will only end up rewarding and entrenching the very same companies that reaped the profits of the invasive surveillance business model that has powered the commercial internet, often at the expense of the public."* (Kak et al., 2023)

Contemporary societies are indeed witnessing an unprecedented concentration of power. The present era might thus be described as "technopolar," to the extent that tech companies wield a new kind of sovereignty (Bremmer and Suleyman, 2023). Furthermore, as suggested by Yanis Varoufakis, the current age could be termed "technofeudalism": a system based on a rentier model mediated by algorithms, which in practice resembles more a feudal scheme than a capitalist one (Monroe, 2023).

5.7 Techno-Solutionism and Techno-Determinism

The representatives of this new power, who, as mentioned earlier, do not exemplify diversity, interpret contemporary times in notably unorthodox ways. Primarily, the worldview of tech leaders, especially those in Silicon Valley, often embraces a strong element of "techno-solutionism," which is the belief that any complex social issue could be overcome with a clear technological solution

(Morozov, 2013), and of techno-determinism, which posits that technology is a key governing force in history.

Another common element is the absolute confidence in the market's power and the primacy of the individual over the collective sphere. Interestingly, these convictions are often augmented by a quasi-religious sense of mission, which involves a somewhat simplistic dichotomy between good and evil. At this point, it suffices to recall Google's motto until 2018, "don't be evil," with the company positioning itself on the "good" side, a notion quite peculiar in a technological context (Assayag, 2024).

5.8 Doomers vs. Accelerationists

Statements from tech leaders and AI scientists provide so many elements for analysis that it would warrant a separate investigation. In any case, it is important to indicate that, within these circles, two seemingly opposing narratives have proliferated: of *doomers* versus *accelerationists*.

In a widely acclaimed 2023 book within the tech community, Mustafa Suleyman compares the evolution of AI to a "wave"—also referred to as "tide," "deluge," "spread," among other terms typically associated with natural disasters (Suleyman, 2023). The choice of words is deliberate, as it suggests that societies cannot but attempt to "contain" the incoming challenges. In the narrative of the book, the source of evil is twofold: it lies in the potential for malicious actors to take control of AI, or for an autonomous AI system to cause a catastrophe.

Regarding the prospect that artificial general intelligence (AGI)—i.e. an intelligence that matches or surpasses human capabilities across a broad range of tasks—could be developed with the potential to dominate or destroy humanity soon, there is no shortage of speculative literature on the topic. Some of the most distinguished AI experts endorse the idea that a "superintelligence"—to use the term coined by University of Oxford philosopher Nick Bostrom (2014)—could pose a global risk. For example, Geoffrey Hinton—a pioneering computer scientist and cognitive psychologist, widely recognised as one of the "godfathers" of deep learning—warns that

> *"[these systems] are more effective at achieving things that are beneficial for us. And they will find it easy to get more power because they will be able to manipulate people."* (as quoted in Wodecki, 2024)

On the other side of the debate, tech businessman and software engineer Marc Andreessen has authored a *Techno Optimist Manifesto* in which he praises the "techno-capital machine" as a force destined to bring about an upward spiral unless hindered by concerns over ethics, safety or sustainability. According to Andreessen, the only option is to accelerate at all costs. Some excerpts from the Manifesto are particularly telling:

> *"We can advance to a far superior way of living, and of being [...].*
> *We believe that there is no material problem—whether created by nature or by technology—that cannot be solved with more technology [...].*
> *We believe in Milton Friedman's observation that human wants and needs are infinite.*
> *We believe in accelerationism—the conscious and deliberate propulsion of technological development—to ensure the fulfilment of the Law of Accelerating Returns. To ensure the techno-capital upward spiral continues forever [...].*
> *We believe technological progress therefore leads to material abundance for everyone.*
> *Becoming Technological Supermen: We believe in nature, but we also believe in overcoming nature. We are not primitives, cowering in fear of the lightning bolt. We are the apex predator; the lightning works for us." (Andreesen, 2023)*

Despite the disputes between doomers and accelerationists over the question whether AGI represents a danger or a boon, for many, this is a misleading discussion. Journalist Cory Doctorow states it quite plainly:

> *"This 'AI debate' is pretty stupid, proceeding as it does from the foregone conclusion that adding compute power and data to the next-word-predictor program will eventually create a conscious being, which will then inevitably become a superbeing. This is a*

> *proposition akin to the idea that if we keep breeding faster and faster horses, we'll get a locomotive."* (Doctorow, 2023b)

Furthermore, for Timnit Gebru—a computer scientist who has made significant contributions to the understanding of the impact of LLMs, which, as seen in Chapter 1, she defined as "stochastic parrots"—, the real issue underlying the controversies (i.e. dangers vs. benefits) surrounding the advent of powerful AGI is far more mundane and pragmatic:

> *"Such claims are incredibly useful for the corporations. Because first of all, look at how our legislators have been distracted by claims of superpower impending apocalypse, impending human extinction. They are not talking about corporations. They are not talking about labor exploitation. They are not talking about data protection or corporations stealing data. They're not talking about corporations not putting in enough resources to ensure that these models are safe [...]. And the second way in which this is extremely helpful to the corporations is that it hypes up the capabilities of the model such that people think that it's going to solve all their problems, and they want to buy it."* (Public Infrastructure, 2023)

5.9 Transhumanism, Effective Altruism and Longtermism

In a recent article, Gebru, with philosopher Émile P. Torres, suggested that the current push to develop AGI is driven by a set of ideologies they term the "TESCREAL bundle", which includes Transhumanism, Extropianism, Singularitarianism, Cosmism, Rationalism, Effective Altruism and Longtermism. They describe the AGI utopia and apocalypse as "two sides of the same coin" and trace the roots of these ideologies back to the Anglo-American eugenics ideology of the early 20th century (Gebru and Torres, 2024). Incidentally, Berkeley professor Judith Butler links the "narcissism that reemerges in the AI dream" with a fascist form of humans (Weil, 2023).

Of the seven ideologies evoked by Gebru and Torres, Transhumanism, Effective Altruism and Longtermism are perhaps the most frequently discussed in the media. *Transhumanism* has been prominent for several decades, advocating the use of technology to enhance human physical and mental capabilities. *Effective Altruism* is a philosophical and social movement led by figures such as Oxford philosopher William MacAskill, which emphasises a utilitarian approach to maximise the global good. It advocates that individuals should aim to do the most good possible in the most rational way, relying on quantitative evidence and analysis rather than emotions or intuitions. *Longtermism*, closely associated with Effective Altruism, emphasises the critical importance of shaping a positive long-term future. Advocates of this ideology argue that if humanity survives the next century, rational calculation suggests prioritising actions that benefit the potentially billions or even trillions of people who—according to these thinkers—will inhabit Earth and beyond in the future.

The principles of Transhumanism are vividly illustrated by Elon Musk's company Neuralink, which aims to merge the brain with AI to "jump-start the next stage of human evolution" (Torres, 2023). Similarly, Altman's donation of 180 million US dollars to Retro Biosciences—a "longevity" start-up focused on rejuvenating bodies—reflects the same ideology (Torres, 2023). Effective Altruism's influence is notably evident among many tech entrepreneurs and start-ups, such as Anthropic (Matthews, 2023). Regarding Longtermism, Musk openly stated in 2022 that McAskill's book *What We Owe the Future*, which champions this perspective, closely aligns with his "own philosophy" (Musk, 2022).

The futuristic dreams of space conquest, inherent in Longtermism and other related ideologies, are also a common theme in discussions among tech leaders. Mentioning Musk, who explicitly aims to colonise Mars with his company SpaceX, is inevitable here. Similarly, Jeff Bezos envisions humans living in giant cylindrical space stations, suggesting there could be "1,000 Mozarts and 1,000 Einsteins" at any given time, and positing that the solar system would be "full of life, intelligence, and energy" (Hart, 2023). Another example is David Holz of Midjourney, who predicts "a billion humanoid robots on Earth in the 2040s and a hundred billion (mostly alien) robots throughout the solar system in the 2060s" (Holz, 2024).

5.10 Science Fiction through the Lens of Big Tech

In addition to ideological influences, tech leaders in the AI era often draw inspiration for their projects from science-fiction novels. For instance, the bot Grok from X is a nod to the sci-fi novel *Stranger in a Strange Land*, where "grok" is a Martian term for "thoroughly understanding something." Furthermore, the brand "DeepMind" pays homage to the Deep Thought supercomputer from Douglas Adams' *The Hitchhiker's Guide to the Galaxy* (Metz et al., 2023b). Moreover, the space utopias envisioned by Musk and Bezos owe much to Iain M. Banks's novel series *The Culture*, which depicts a post-scarcity society in which humans live under the protection of vast, sentient intelligences and are freed from material needs.

However, it is striking that the reading tech titans make of these works of fiction is partial, as they borrow certain words or images but miss their deeper meaning. For example, the appreciation that figures like Musk, Bezos and others have for Banks's utopian sci-fi novels aligns with their technological visions but completely overlooks his harsh critiques of wealth concentration and market-driven societies, leading some to lament that they simply did not understand the books (Power, 2022).

5.11 Power and Discourse in the Age of AI

Or perhaps the issue lies in their utilitarian reading of these texts—they selectively extract aspects that serve their personal or corporate purposes, similar to their approach to the aforementioned philosophies. It is important to remember that one of the champions of Effective Altruism was Sam Bankman-Fried, the former CEO of FTX—who advocated for amassing as much money as possible, ostensibly to donate it later. Eventually, he was convicted on charges of wire fraud and money laundering. This situation casted a shadow on the potentially manipulative exploitation of those doctrines, clearly illustrating how they can be used to mask unethical practices (Christian, 2023).

Indeed, behind the narratives and metaphors presented by AI leaders, it is crucial to closely observe their actual behaviours. The events at OpenAI in 2023 are a

telling example. On November 17, Altman was fired by the board of directors because he was not "consistently candid in his communications with the board" (Metz et al., 2023a). Altman's firing sparked a crisis, with nearly all 800 OpenAI employees threatening to leave for Microsoft with him and Greg Brockman, who resigned as president and chairman in solidarity. Notably, on November 21, Altman was reinstated as CEO, removing those who opposed him from the board, including Ilya Sutskever. The episode surprised much of the specialised press due to the lack of precise information on the occurrence. The truth is that a tool like ChatGPT, having become almost indispensable to tens of millions of individuals and thousands of businesses across the globe, seemed suddenly in jeopardy. While, just a few months earlier, OpenAI had been obsessively fixated with issues of technical safety and the existential risk posed by AGI (Edwards, 2023), it became evident to the public that the real threat originated from a different source: the conflict of human interests, unfolding amidst deep-seated opacity.

6. The Geopolitics of AI: A Global Chessboard in Contest

6.1 "Digital Empires"

If major technology companies represent a new type of power, it is worth acknowledging that these entities operate within nations or regional blocks, each with its own challenges and priorities, highlighting the central role of geopolitical considerations. On a global scale, there is a remarkable variety of national perspectives on AI—indeed, each country or region presents its own sociotechnical imaginaries in relation to AI strategies (Bareis and Katzenbach, 2021).

In any case, it could be posited that three major entities currently dominate: the US, China and the European Union (EU). According to the leading scholar on EU regulatory power and digital regulation, Professor Anu Bradford, these constitute distinct "digital empires," each following a unique rationale: the US is market-driven, China is state-driven, while the EU is rights-driven (Bradford, 2023).

6.2 The United States

The case of the US is characterised by significant strengths, including a robust innovation ecosystem, substantial investments in AI research and development, leading universities as well as globally dominant tech giants and AI start-ups, particularly those based in Silicon Valley, as detailed in Chapter 5. Although the predominant approach can be defined as market-driven, the reality is much more complex. Various tech companies maintain close collaborations with different state agencies, such as the partnership between OpenAI and the Pentagon on cybersecurity tools (Hammer, 2024). In fact, there is no shortage of voices calling for an increased public sector involvement in AI investment, partially inspired by the investment strategies of China and the European Union, to avoid leaving the field solely to private corporations (Schneier and Sanders, 2023).

The leadership of US big tech in NLP, particularly in the development of multilingual LLMs, constitutes a clear geopolitical advantage. As AI agents proliferate, the ability of these systems to communicate in various global languages will become a privileged asset, one that few outside the US will possess.

Geopolitical competition remains a constant item on the American AI agenda and was heightened when Eric Schmidt in 2017 warned that China had declared its intention to become the primary AI innovation centre by 2030 and "dominate the industries of AI" (Shead, 2017).

In the context of US AI regulations, several regulatory frameworks had already been established prior to 2023, including the "Blueprint for an AI Bill of Rights" from October 2022. This blueprint outlined five essential principles designed to protect the American people's rights: the creation of safe and effective systems, protections against algorithmic discrimination, data privacy, the provision of notice and explanation, and the assurance of human alternatives, consideration and fallback mechanisms (The White House, 2022).

The "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence", issued by President Biden in October 2023, furthered these efforts by setting new, comprehensive standards for AI safety and security, while also imposing rigorous requirements for AI transparency, testing, red-teaming and cybersecurity (The White House, 2023). The Executive Order also catalysed a significant coordinated programme across the US government to enhance American privacy, promote equity and civil rights, support consumers and workers, stimulate innovation and competition as well as bolster American global leadership, to name just a few goals. Additionally, it introduced notification requirements for foundational models as a precursor to a potential licensing system for high-risk models (Sun, 2023).

According to US Vice President Kamala Harris, AI systems cannot be developed with a focus solely on economic benefits without considering their social impacts:

> *"We reject the false choice that suggests we can either protect the public or advance innovation. We can—and we must—do both. And we must do so swiftly, as this technology rapidly advances […]. History has shown [that] in the absence of regulation and strong government oversight, some technology companies choose to prioritise profit over the wellbeing of their customers, the security of our communities and the stability of our democracies."* (as quoted in Wintour, 2023)

The Executive Order reflects a strong emphasis on promoting competition and preventing market dominance, as seen in its section 5.3(a). Reflecting this focus, the Federal Trade Commission (FTC), under Lina Kahn's chairmanship, has been closely monitoring acquisitions, investments and partnerships in the AI industry to pre-empt potential vertical integrations and the consolidation of dominant market positions. Kahn has stated that the goal is to prevent the formation of monopolies before they become entrenched (Heath and Fischer, 2024).

Furthermore, following the implementation of the Executive Order, the US government has announced the establishment of an AI Safety Institute Consortium (AISIC) and an AI Safety and Security Board that includes some academic experts but primarily industry leaders such as Sam Altman and Dario Amodei. This move has not been without its critics: Timnit Gebru, for example, employed the metaphor "foxes guarding the hen house" to criticise the board's composition (Gebru, 2024).

In any case, it remains to be seen what aspects of the Executive Order will be retained after a potential Republican victory. Indeed, candidate Donald Trump has already warned the public that, were he re-elected, he would repeal the regulation (MAGA War Room, 2023).

6.3 China

In China, the advantages are also significant, including vast data pools arising from a market of over a billion connected users, extremely dynamic domestic tech giants, rapid adoption of AI across various sectors, active government engagement in the overall strategic direction and close collaboration between the government and tech companies. China's technological strength is undeniable: according to a 2023 policy brief by the Australian Strategic Policy Institute (ASPI), China leads in 37 out of 44 areas covering crucial technologies such as defence, space, AI and robotics. While the US holds the second place in most categories, China has pulled significantly ahead through its exceptional research and years of strategic policy work aimed at cultivating tech talent and investments (Gaida et al., 2023).

China undoubtedly seeks to position itself as a prominent player in the field of AI. However, it is necessary to note that in Eric Schmidt's interpretation, a translation error may have occurred. Indeed, the original Chinese plan referred to the intention of becoming "a primary AI innovation centre of the world by 2030," not "the world's primary AI innovation centre" (ÓhÉigeartaigh et al., 2020: 585). This incident may highlight the misunderstandings that the US and the West in general have regarding the Chinese context. The linguistic imbalances in favour of English, noted in Chapter 2, may have played a significant role here. While Chinese researchers generally speak English—allowing them to access scientific production and regulatory documents in said language—the reverse is not the case: the English-speaking community has far less likelihood of accessing original texts written in Chinese, which limits their understanding of Chinese advancements (ÓhÉigeartaigh et al., 2020).

China has established a comprehensive framework of AI regulations, notably including the "Algorithmic Recommendation Regulation", which took effect in March 2022. This regulation targets the algorithmic dissemination of online content, requiring all models to be filed in China's "Algorithm Registration System" (Government of the People's Republic of China, 2022). Following this, the "Deep Synthesis Regulation" was enacted in January 2023, focusing on AI-based technologies that generate synthetic content like images, audio, text and deepfakes. It demands that all models be registered, content be labelled and dissemination of prohibited information be prevented (Cyberspace Administration of China, 2022). The "Administrative Measures for Generative AI Services", implemented in August 2023, stipulate that AI models should undergo security assessments before launching services, label AI-generated content and ensure the use of screened training data (Cyberspace Administration of China, 2023). Finally, the "Draft Basic Security Requirements for Generative AI Services," introduced in October 2023, focuses on the legality of training data, establishing IP management strategies and ensuring the generation of accurate content from registered models (National Information Security Standardisation Technical Committee, 2023).

New regulations emerge monthly, and as described in Chapter 4, they typically include measures to censor or ensure control over both the input (the training data) and the output (both AI-generated results or content selections made by

AI algorithms) in conjunction with primarily local private sector stakeholders. It is important to note that ChatGPT is blocked in China.

Another typical feature of the Chinese regulatory approach is its iterative character, based on learning-by-doing and focusing on regulations that generally target specific applications. This strategy's pragmatism and effectiveness have drawn praise even from experts in the US (Sheehan, 2023).

6.4 The European Union

The EU, meanwhile, boasts world-class universities and institutes conducting cutting-edge AI research, a single market of over 500 million people and funding programmes such as Horizon Europe, which provide significant support for research and development, and supercomputing infrastructure. In response to American and Chinese pre-eminence, many voices in Europe view open-source (Robertson, 2023) and niche innovations as opportunities (Mhalla, 2023).

In regulatory terms, the EU possesses some of the most comprehensive and ambitious frameworks in the world. Articulating with other regulations such as the General Data Protection Regulation (GDPR), the Data Act, the Digital Markets Act (DMA) and the Digital Services Act (DSA), the AI Act was approved in March 2024 (European Parliament, 2024). It addresses a range of risks and establishes clear responsibilities for both public and private actors within and beyond the EU. It applies extraterritorially, meaning it affects any AI system marketed inside the Union or whose use impacts individuals within the EU.

The Act classifies AI risks into four distinct groups: those with minimal risk, limited risk, high risk and unacceptable risk. For generative AI, it mandates clear transparency and disclosure protocols. General-purpose AI and foundational models are subject to strict rules regarding technical documentation, training data, summaries and safeguards for both copyright and intellectual property rights. Furthermore, the Act introduces systemic risk assessments for models with significant computational power, with a threshold of 10^{25} FLOPs—capturing the currently most advanced general-purpose AI models, namely GPT-4 and Gemini. High-risk AI systems must be trained and tested using

sufficiently representative datasets to minimise the risk of unfair biases being embedded in the model, and there are specific exemptions for national security, military and defence purposes. The AI Act also establishes the formation of an AI Board, an AI Office, an Advisory Forum and a Scientific Panel of independent experts.

It is worth noting that the initial drafts of the AI Act began negotiations in 2021—well before the surge in generative AI and the emergence of systems like ChatGPT—, which complicated the original framework based solely on different levels of risk. Consequently, the category of "general-purpose AI" was added. This addition faced significant criticism from several representatives of the European AI industry, who argued that specific applications, not general models should be regulated (Wanat, 2023). Be it as it may, the full implementation of the AI Act is not immediate, with its complete rollout scheduled for 2026.

The EU AI Act is poised to exert a substantial global influence, similar to the impact seen with the GDPR, showcasing the EU's ability to shape the international regulatory environment through the so-called "Brussels Effect." By establishing rigorous regulations for AI, the EU is likely to not only affect product offerings outside its borders—a de facto influence—but may also prompt other jurisdictions to adopt similar regulations—a de jure influence (Siegmann and Anderljung, 2022).

Within the EU, France, Germany, Spain and Italy stand out for their national plans, investments and start-ups in the AI sector. Specifically, France's recent report *AI: Our Ambition for France* proposes investing 5 billion euros annually over the next five years in various strategic AI areas (Artificial Intelligence Commission, 2024: 4).

6.5 Other Relevant Players

On the broader international stage, countries such as the UK, Japan, South Korea and Canada also distinguish themselves with leading universities, advanced research institutes, innovative tech start-ups and robust investment in AI initiatives. However, these countries have not fully adopted specific

regulatory frameworks for AI, with projects either in draft stages or undergoing legislative development and not yet enacted into law (Milligan, 2024; Sakai et al., 2024; Kyoung Ko, 2024; Bath, 2024).

In November 2023, the UK government hosted the first global AI Safety Summit at Bletchley Park, a site steeped in historical significance as the World War II codebreaking centre and a birthplace of modern computing. The event saw participation from 28 countries and concluded with a declaration advocating for international collaboration to address AI risks ("The Bletchley Declaration", 2023). However, the summit's lack of focus on other significant challenges such as AI's environmental impact or issues of bias and discrimination, along with the dominant presence of big tech companies, raised concerns about the scope and priorities of such initiative (Macaulay, 2023).

6.6 The Global South

While the gap in AI capabilities between the US, China and other developed countries like France, Germany, the UK and Japan is significant, the disparity with countries in the Global South is even larger. Apart from some Gulf countries like the United Arab Emirates—which invest heavily in chip production and the creation of open-source systems such as the Falcon model (Perrigo, 2024)—, it is perhaps India, bolstered by its vast domestic market and its prominence in IT research, that stands out as one of the few Global South countries able to devise a comprehensive national AI strategy. This aspiration is evident in the *India AI 2023* report, compiled by the Ministry of Electronics and Information Technology (MeitY, 2023). India is also investing in building its AI infrastructure as part of its digital public infrastructure (DPI) plan. According to the Minister of State for Electronics and IT, Rajeev Chandrasekhar, this initiative is crucial for the nation's future:

> *"We are determined that we must have our own sovereign AI. We can take two options. One is to say, as long as there is an AI ecosystem in India whether that is driven by Google, Meta, Indian startups, and Indian companies, we should be happy about it. But we certainly don't think that is enough."* (Grover, 2023)

This perspective is indeed shared by Jensen Huang, CEO of Nvidia, who believes that every country needs to have its own AI infrastructure to leverage economic potential while protecting its own culture:

> *"There are some interests to scare people about this new technology, to mystify this technology, to encourage other people to not do anything about that technology and rely on them to do it. And I think that's a mistake."* (Reuters, 2024)

Nevertheless, many countries in regions such as Africa and Latin America lack this scale and investment capacity, and face enormous infrastructural problems. This explains why, despite the presented opportunities, few of these nations have outlined their own AI agendas (Okolo, 2023). In the AI Readiness Index proposed by Oxford Insights (2023), which considers variables related to governance, technology, data and infrastructure, no African country is among the top 60 in the world. Moreover, 21 out of the 25 lowest scoring countries are on that continent. These findings are consistent with the report *AI Governance for Africa*, published by the Thomson Reuters Foundation in November 2023, indicating that only 7 of the 55 African Union countries have implemented a national AI strategy (Davis et al., 2023: 12).

The serious obstacles that hinder the formation of a robust and sustainable AI ecosystem in the Global South do not imply the absence of highly dynamic and engaging initiatives. Competitive AI start-ups like Instadeep, which emerged in Tunisia, or cutting-edge research centres such as the African Research Centre for Artificial Intelligence (ARCAI) in Congo (UNECA, 2022) are just a few examples of the diversity flourishing in developing regions. However, without a coordinated national or broader agenda, the countries of the Global South could be perpetually relegated to the role of simple suppliers in an international division of labour that would seriously harm the prospects of these nations, potentially turning them into mere "AI vassal states" (The Economist, 2018).

Indeed, one of the most apparent risks for developing countries in the AI era is brain drain. The availability of better academic and professional opportunities, advanced research facilities and a more substantial funding in developed countries naturally attracts top talents to leave their home countries. Moreover, obtaining visas for researchers from the Global South to attend key sector

events—which predominantly occur in the US, Canada or European countries—is often a complicated process. This barrier creates significant asymmetries, disadvantaging these researchers compared to their international peers and limiting their exposure to global networks.

On another note, many developing countries are targeted for outsourcing low-paid, tedious and often psychologically taxing tasks. For instance, OpenAI utilised an outsourcing firm in Kenya to obtain the necessary data labels for its models. This firm processed tens of thousands of snippets of text, some of which were sourced from the darker parts of the internet and included descriptions of extreme and disturbing scenarios, ultimately resulting in trauma for many of the workers involved (Perrigo, 2023b).

Many countries in the Global South have attempted to emulate the success of their counterparts in the Global North by forcibly implanting similar technological approaches. However, this imitation has often resulted in oversimplified responses to complex social issues. Such an approach not only overlooks the nuanced needs and agency of individuals but also risks perpetuating a colonial-like exploitation of data and resources (Birhane, 2020). In this context, there is a growing call for a "decolonial AI," aimed at empowering actors in the Global South (Hao, 2022).

6.7 International Cooperation

In a context of deep global asymmetries, there are countless projects promoting an international scientific and technological cooperation. Indeed, the development of AI is particularly conducive to cooperation, as much of its progress is made by an already highly internationalised academic community (ÓhÉigeartaigh et al., 2020).

China, for example, announced an ambitious cooperation plan called the Global AI Governance Initiative (GAIGI), designed for countries of the Belt and Road Initiative (Cong and Yeping, 2023). The US, on the other hand, has established recent partnerships with Japan and the EU in AI and semiconductors, among other areas (Ogura and Liu, 2024; Greenacre, 2024).

However, it seems clear that many of these cooperation programmes follow a geopolitical logic that could lead to a kind of "AI Cold War" between the US and its allies on one side and China and its partners on the other, as acknowledged by the participants themselves (Walla, 2024). In this rivalry, both giants may attempt to block each other's access to key resources—hardware, rare minerals, data and more (Márquez Lartigue, 2023b).

Despite these competitive dynamics between major global players, it remains essential not to overlook collaborative projects implemented beyond this rivalry. In February 2024, the African Union Development Agency released a policy draft detailing a continental AI regulatory framework, which includes industry-specific codes, standards and certification bodies, as well as national AI councils for oversight. Such initiatives can facilitate a possible continental AI strategy (Tsanni, 2024). Given the challenges in devising AI strategies and regulations in developing countries, this type of regional work may prove indispensable.

6.8 Global Frameworks

Beyond regional agreements, globally developed guidelines—especially those under the auspices of the United Nations (UN)—are of particular importance. In November 2021, all 193 Member States of UNESCO adopted the *Recommendation on the Ethics of AI*. While previous documents regarding AI ethics had been proposed by entities such as the Organisation for Economic Co-operation and Development (OECD), the G7 and others, this UNESCO document stands as the first-ever global standard in its field (UNESCO, 2021).

The *Recommendation* delineates a framework of values aimed at promoting and protecting human rights, dignity and the environment. It advocates for critical principles such as transparency, accountability and an adherence to the rule of law. Moreover, it proposes specific policy measures addressing data governance, international cooperation, environmental concerns, gender, culture, education, the economy and labour, health, among others. Notably, with AI and culture, UNESCO's *Recommendation* is one of the few—if not the only—international instruments dedicating a specific chapter to this topic. It was drafted by an independent committee of experts from 24 countries across the Global North

and the Global South and received extensive input from specialists and organisations worldwide.

In November 2023, during the UK's AI Safety Summit, UN Secretary-General António Guterres called for a united global strategy to "get ahead of the wave" and tackle the challenges of AI safety and ethics. Around the same time, the UN established an Advisory Body on AI to examine models of technology governance, comprising 38 international experts from governments, civil society, academia and the tech industry—indeed, the committee includes representatives from Google, OpenAI, Microsoft and Sony (UN, 2023). In March 2024, drawing from UNESCO's *Recommendation* and other relevant documents, the UN General Assembly adopted a US-proposed resolution on promoting "safe, secure, and trustworthy" AI (UN, 2024).

6.9 Tech Diplomacy

In addition to developments in each country and the interactions between states, it is important to highlight the area of "tech diplomacy," namely, the engagements between states and major tech companies. As the influence of tech titans has expanded—as discussed in Chapter 5 under the concept of a "technopolar era"—, various countries have also pursued dialogues with entities such as Facebook/Meta, Google/Alphabet and other Silicon Valley-based tech companies. Denmark was the first to establish the position of a tech ambassador in 2017, a trend that was later followed by countries including Austria, France, the UK and Brazil.

A 2023 report found that there are currently 63 active government institutions with a presence in the San Francisco Bay Area, predominantly from European countries (Mind the Bridge, 2023). However, countries from the Global South are underrepresented in this group. On the agenda, AI continues to be a top priority, underscoring its cross-sectoral impact.

In February 2023, the Tech Diplomacy Network was launched in San Francisco. This initiative is a collaboration between the Berggruen Institute, the World Economic Forum (Centre for the Fourth Industrial Revolution/C4IR), the Bay Area Council Economic Institute and the DiploFoundation. Its goal is to

enhance collaboration and dialogue between the diplomatic community, civil society and the tech industry both in the Bay Area and beyond, creating a platform for shared understanding and strategic alignment in the rapidly evolving technological landscape (Ittelson and Rauchbauer, 2023).

7. Working in International Cultural Relations in the Age of AI

In the increasingly complex technological and geopolitical landscape described in the previous chapters, outlining programmes and long-term strategies is not always so straightforward for institutions working in ICR. The use of AI as a tool by these institutions also poses several significant challenges.[17] The following sections will analyse the most pressing issues in these areas, primarily drawing from the interviews conducted with specialists active in ICR.[18]

7.1 Need for Information, Skills, Collaborations and Impact

First, there is a need for accessible information on AI for ICR actors. Several individuals interviewed for this research noted that the technological landscape is so dynamic and fast evolving that it is exceedingly demanding to keep pace with the rapid, also weekly changes. Without accessible forums or sources of updated information useful for those outside of computer science or the entrepreneurial sphere, there is a need for something akin to an "idea buffet" or a resource to draw from when thinking about new initiatives. Additionally, there often are insufficient resources to sustain dedicated staff in the digital and AI domains or to train the current team on emerging technology issues.

Second, a recurrent topic is the divide between the tech sectors on one side and the political and cultural sectors on the other. As one interviewee stated, "We need to launch new [intersectoral] collaborations; we can no longer act alone!" Along the same lines, a specialist indicated that today, more than ever, it is crucial to understand the notion of "cultural diversity" in precisely "culturally diverse" ways: indeed, there is currently a risk of isolationism when the concept of "cultural diversity" is taken to mean "my values are my values," which could stifle the conversation between institutions from different countries at a time when more synergies are essential to comprehend the global impacts of AI.

[17] The distinction between AI as a tool, a topic or a context-changing factor has already been addressed by Höne et al. (2019) in their analysis of AI and diplomacy, and it may also be relevant for studying the impact of AI on ICR.

[18] The list of interviewees can be found at the end of this report.

Furthermore, another problem is how to think about the issue of collaboration and the concept of the "other," when activities are mediated by non-organic entities that are also not free from biases.

Another interviewee emphasised the increasing need to promote face-to-face meetings and activities today, as there is a public fatigue with digital-only interactions in this post-Covid era. However, the logistics and costs of assembling a face-to-face panel of guests from different countries can be demanding.

Third, the lack of measurability of ICR actions poses a challenge in initiating collaborations. An interviewee explained that, to establish long-term collaborations with public funding agencies, it is necessary to demonstrate impact metrics for proposed ICR initiatives. However, in areas such as cultural diversity or freedom of expression in the age of AI, the effects of any programme are often long-term and not always easy to quantify.

Finally, the lack of political influence of sectors working in ICR has also been noted as a challenge. Indeed, the voice of tech actors often plays a more decisive role in shaping national AI plans or in AI regulatory frameworks than that of cultural actors or social organisations. As one interviewee remarked, there is generally a discourse of AI as "something that the arts and culture sector should respond to, letting tech people lead."

7.2 AI as a Tool for ICR

While current AI applications like ChatGPT and others are user-friendly and could contribute to greater productivity, assist with translations and even provide inspiration for new programmes, these systems can produce outputs that are problematic for work in ICR. One interviewee mentioned the case of Gemini, which, when asked to draw a Curupira—a character from Tupí-Guaraní mythology—, ended up depicting a tiger, demonstrating its limited understanding of non-Western cultures. Using such tools for intercultural communication can lead to a deceptive perception of mutual comprehension or, conversely, might engender artificial barriers to understanding or even provoke resentment, stemming from the biases and errors ingrained in the system.

In practice, automation cannot replace the interpersonal relationships and informal conversations that are fundamental to international dialogue (Márquez Lartigue, 2023a). Unlike humans, machines do not grasp the actual meaning of conversations and lack the basic soft skills essential for ICR, such as cultural sensitivity and empathy (Garcia, 2024). Furthermore, devising programmes at the intersection of AI and ICR necessitates innovative and original perspectives. A cultural agency representative lamented that ChatGPT was actually quite uninspired in generating ideas in this field, possibly because it was trained on data from the past.

Another issue frequently mentioned is that, when it comes to using more sophisticated AI tools, the responsible parties in an institution who manage or develop these tools are often part of a separate IT team and are not always aware of the cultural or social issues involved in their use. Privacy and confidentiality concerns in the use of third-party AI tools were also highlighted.

7.3 Narratives, Metaphors and Imaginaries

According to some interviewees, the exploration of local imaginaries and narratives—both contemporary and historical—around AI is an area that could be developed more deeply. However, it is often overlooked due to the prevailing belief within the tech community that there is only one valid approach to this subject, which explicitly rejects any consideration of local contexts. The lack of intercultural sensitivity in many tech companies may lead them to view the differences inherent in diversity as a bug to be fixed, thus potentially eliminating one of the most interesting aspects of work in ICR.

However, as one expert interviewee noted, the "future will always be shaped by narratives; we cannot live without them." Indeed, there is no single way to think about the topic of automation, and the weight of the cultural dimension is considerable. Ultimately, the narratives and metaphors employed—consciously or not—to refer to these issues will impact the future development of technology and its regulation (Maas, 2023; Floridi and Nobre, 2024).

An interviewee emphasised the need to challenge Silicon Valley and to tell them that they "are not as innovative as they think!" Another interviewee pointed out that, in the current context, they were more concerned with "human stupidity than "artificial intelligence." This is particularly applicable in cases in which general users deploy AI systems uncritically, believing that the machine would resolve everything on its own.

7.4 Projects at the Intersection of AI and Cultural Relations

Despite the difficulties, many institutions involved in ICR have already launched a number of successful initiatives related to AI. These initiatives include organising debates on cultural and social impacts of AI, raising awareness of recent advancements, supporting the (inter)national projection of cultural and creative industries in the AI era, and developing programmes to promote diversity in the AI sphere.

The Goethe-Institut, for instance, conducted the "Generation A=Algorithm"[19] project from January 2020 to December 2021. This project expanded the conversation on AI from specialised experts to the general public. It aimed to promote cross-border and interdisciplinary dialogues that critically assessed AI technologies and involved participants in developing algorithmic systems. The initiative primarily sought to enhance young adults' understanding of AI, equipping them with tools to influence their daily interactions with algorithms as part of the next generation—Generation A.

In 2023, the Goethe-Institut also launched the "A2Amplify" project[20] to address the lack of diversity in AI development teams, which often results in technologies that fail to meet the needs of or even harm marginalised groups. The initiative aimed to include voices from the Global South to help narrow the digital divide exacerbated by the predominance of machine-learning datasets from the Global North. It focused on empowering coders from these regions, integrating their interests into global AI debates, and ensuring that AI

[19] https://www.goethe.de/prj/zei/en/prj/22175938.html.
[20] https://www.goethe.de/prj/aia/en.

applications reflected the cultural realities, needs and interests of the Global South.

The Institut Français (IF) offers several programmes related to the digital transformation, prominently featuring AI as a key topic. In November of each year, the IF hosts *Novembre Numérique*[21] ("Digital November"), an international celebration of digital cultures. The event, which was on its seventh iteration in 2023, aims to showcase innovative projects and foster discussions around key digital themes, including AI.

Moreover, the IF's "Creating with Artificial Intelligence" series showcases artists who utilise AI as a fertile ground for creativity. This collection of activities highlights how AI is not only a tool for generating unique artistic works but also a medium for reflecting on our digital world and highlighting its potential misuses. It features a range of projects that explore themes such as the relationship between humans and nature, the embodiment of technology in humanoid forms and the ethical concerns over AI.

In 2020, the EU National Institutes for Culture (EUNIC) organised a comprehensive programme of debates on "how digital transformation and artificial intelligence challenge cultural relations," during the EUNIC DX/AI Week.[22] The following year, EUNIC also held multiple "AI Science Cafes" to discuss topics at the crossing of AI and art, ethics, cybersecurity and more.

Furthermore, in 2023, the project "AIIA—Indigenous Appropriation of Artificial Intelligence,"[23] supported by the IberCultura Viva cooperation programme, brought together nine indigenous communities from Argentina, Brazil and Chile. They collaborated on creating contemporary digital artworks using AI tools, and it was aspired to empower local communities through the critical use of these technologies.

[21] https://www.institutfrancais.com/fr/offre/novembre-numerique-2023.

[22] https://www.eunicglobal.eu/news/eunic-dx-ai-week.

[23] https://iberculturaviva.org/identidad-afro-fe-y-fiesta-arte-indigena-e-inteligencia-artificial-los-proyectos-de-brasil-seleccionados-en-la-convocatoria-de-apoyo-a-redes-2023/?lang=es.

7.5 Art + Tech

In June 2019, European cultural institutes from Austria, France, Germany, Italy and Switzerland established the EUNIC Silicon Valley Cluster to jointly explore the junction of art and technology. The initiative aimed to review existing partnership models between artists and technologists, intending to map out the full potential of these interactions. Funded by the EU, the project "The Grid"[24] was launched to disseminate findings from investigating possible models of collaboration between artists, cultural institutions and the tech industry.

Notably, The Grid seemed to take an inverse approach compared to previous initiatives; it focused less on integrating AI or tech into the cultural sector or ICR and more on bringing artistic perspectives to the tech universe, leveraging the robust networks built by tech diplomacy. A key driver of The Grid was Open Austria,[25] the Austrian diplomatic representation in Silicon Valley, which emphasised the significance of the arts in "redefining what it means to be human in the age of AI" and highlighted the role of artists "as futurists and storytellers who are uniquely positioned to identify and articulate both the opportunities and challenges of frontier technologies" (Open Austria, 2021).

During the two-year lifespan of The Grid, two reports titled *Art + Tech* were produced, with the 2019 edition exploring collaboration models between technology companies and the arts in the San Francisco Bay Area (Hochman and Reben, 2019). The analysis, derived from interviews and an online survey, highlights the creativity of existing partnerships and identifies potential roles for the EU in fostering deeper connections between these sectors. Key initiatives include artist residencies at companies like Facebook and Amazon, which emphasise the unique role of artists in examining the broader implications of technology. The report portrays artists as crucial in prompting companies and society to think deeply about the consequences of digital advancements, focusing on identifying needs rather than finding "solutions."

The 2020 edition of the report examines the forces that have influenced interactions between art, technology, counterculture and industry in the Bay

[24] https://www.eunicglobal.eu/projects/european-houses-of-culture-usa-the-grid
[25] https://www.open-austria.com

Area (Chang, 2020). It underscores how the tech sector's priorities shape the field. The report notes an extractive relationship between artists and the tech industry, further exacerbated by the challenges in art funding due to a prevailing shift towards Effective Altruism, where the focus is on quantifiable impacts and returns on investment.

7.6 The Dawn of Digital Humanism

In the age of AI, the advances made by diplomats, researchers and artists at the intersection of arts and technology offer invaluable lessons for work in ICR. In a 2022 paper titled "How to Be a Digital Humanist in International Relations: Cultural Tech Diplomacy Challenges Silicon Valley," Carla Blume and Martin Rauchbauer—formerly President of EUNIC Silicon Valley and Head of Open Austria, respectively—explore different facets of the emerging field of Digital Humanism, which may prove critical in a world "radically transformed by artificial intelligence and biotechnology" (Blume and Rauchbauer, 2022).

It is worth noting that the concept of "Digital Humanism" has gained significant momentum since the release of the *Vienna Manifesto on Digital Humanism* (Werthner et al., 2019). Written in May 2019 and endorsed by over 1000 global leaders, the manifesto calls for a re-assessment of technology's impact on society. It champions the development of digital technologies that reinforce human values and support democratic, equitable societies.

Along these lines, Blume and Rauchbauer advocate for a conversation among diverse stakeholders from the cultural, political and technological realms. The authors acknowledge that this effort is not without its challenges, partly because the various participants often speak different "languages." For instance, when a Silicon Valley entrepreneur discusses "technology," "truth" or "freedom", they do not necessarily ascribe the same meanings to these terms as a policy-maker might. This discrepancy arises in part because these stakeholders hold divergent views of humanity. Tech leaders' mindset—although they are usually unaware of this—is deeply rooted in the fervour of the historic Gold Rush, which aligns with the American dream of global exceptionalism and the urge to push the Western frontier. The doctrine of absolute freedom of information and resistance to centralised control, championed by many internet pioneers, stems

from the hippie counterculture that emerged in San Francisco during the 1960s. Alongside significant technological and scientific breakthroughs, these traditions have merged into Transhumanism and other ideologies that, in practice, support an elite whose major projects include ambitions for humanity to transcend the confines of planet Earth.

These differences in language and—more profoundly—in perspectives serve to hinder the necessary collaborations among various groups, exacerbating the disconnect. This lack of communication makes the situation worse both for tech companies, which miss out on new, possibly beneficial artistic and philosophical insights, and for the wider world, which sees its future being shaped "by a handful of individuals with a ticket to Mars" (Blume and Rauchbauer, 2022: 103). Therefore, according to the authors, there is a clear need today for an approach based on Digital Humanism, "equipped with a certain pragmatism that is tilting neither toward delusional techno-utopianism nor fatalistic fear-mongering."

8. Conclusion and Recommendations

As this research has attempted to show, the integration of AI systems introduces both opportunities and challenges in the most critical action areas for ICR. Indeed, there are hardly any aspects of ICR that remain untouched by this technology that is so deeply entwined with cultural dynamics.

From a technical standpoint, AI systems enable the execution of an increasing number of complex tasks. However, they may suffer from issues such as the so-called "hallucinations," the potential for a model collapse due to a lack of original content, embedded biases, value conflicts and security vulnerabilities, among other issues.

Regarding languages, even though AI systems can help with areas such as translation, language learning and the preservation of linguistic heritage, there are enormous imbalances between languages—particularly between English and all others—as well as challenges that could threaten the very survival of low-resource languages.

In the creative sectors, the impact of AI is undeniable. Integrating this technology into literature, visual arts, cinema, museums, video games and others offers expanded artistic possibilities, enhanced productivity for cultural and creative industries as well as a significantly enriched content offering. However, the lack of transparency in using artworks to train AI systems, combined with the potential displacement of artists and entrepreneurs along the cultural value chain, could lead to a rapid degradation of the creative ecosystem.

In the information environment, AI tools can potentially improve communication and streamline content moderation. Nonetheless, challenges in this area are escalating daily: there is a proliferation of synthetic content, misinformation, disinformation and AI-driven manipulation. Simultaneously, the strategies employed by search engines, social networks and other platforms can lead to a deterioration of the situation. In addition, censorship methods in the age of AI are becoming more sophisticated.

Regarding technological actors, although new AI start-ups emerge on a monthly basis, the few that survive eventually align with major tech corporations in a sector marked by a conspicuous lack of diversity. In this landscape, risks such as economic concentration, job loss and growing inequality are expected to magnify in the future. Furthermore, the ideologies of tech leaders often take on

rather striking forms, while the professed commitments of many of them frequently contradict their actual practices.

Finally, international relations in the AI era are not free from significant geopolitical tensions, particularly due to the escalating rivalry between the USA and China. Globally, AI strategies and regulations are proliferating, though relatively few states have currently managed to design and adopt long-term policies. Particularly countries in the Global South are at risk of becoming subjected to a colonial-like exploitation of data and resources. Despite these challenges, several global frameworks have emerged under the auspices of entities such as UNESCO and the broader UN system, along with a growing number of tech diplomacy initiatives that contribute to a closer engagement between the various stakeholders.

In this scenario, the work for ICR institutions can be very complex. As expressed by the protagonists, there is a need for more data, skills and collaborations with the tech sector. Simultaneously, the realm of imaginaries and narratives surrounding AI represents an untapped universe with enormous potential for influence on the development and regulation of AI. Several projects have been initiated at the junction of ICR and AI, though much remains to be done, especially to promote a new type of approach that could be described as "Digital Humanism."

In the current context, despite all the challenges, the work in ICR in the AI era has become more crucial than ever. Indeed, the research, application and oversight of these powerful systems should not be left solely in the hands of the tech sector or governmental agencies predominantly focused on the "AI race." ICR institutions can take on an active role, ensuring the ethical use of AI technologies and getting deeply involved in reshaping the imaginaries and narratives that ultimately guide most developments and regulations of AI on a global scale.

In light of these needs, the following section presents a series of recommendations grouped into different thematic areas aimed at actors working either directly or indirectly in ICR, with a particular focus on ICR institutions. Each of these proposals can serve as a foundation for comprehensive programmes.

8.1 Training and Awareness

- Through courses and conferences, promote AI and data literacy among the internal teams in the institution as well as among representatives from government agencies, other cultural organisations, cultural and creative industries, artists and civil society.
- In these trainings, include topics related to the main current AI applications, underlying technologies, advantages, technical and social challenges as well as main actors involved.
- Develop new, accessible forms of educational dissemination on AI for laypersons, to be published periodically.
- Encourage lifelong learning and highlight the importance of technical reskilling and upskilling in sectors linked to ICR.
- Foster cultural awareness for the actors involved in the development, use and regulation of AI technologies. This includes, among other significant issues, paying attention to challenges such as cultural biases and the narratives involved.
- Contribute to greater empowerment of citizens to identify and counteract the detrimental effects of AI technologies, both in the Global North and the Global South.
- Invest heavily in soft skills that are indispensable to the field of ICR in the AI era, such as cross-cultural communication, empathy, adaptability and negotiation.
- Enhance internal capacities for data use and protection as well as for developing in-house AI tools.
- Invest in foresight to anticipate changes in the tech landscape and act as early as possible.
- Encourage the translation and publication of key texts on current AI trends in languages other than English.

8.2 Collaborations

- Link up initiatives among ICR institutions that share similar challenges around the world, contributing to the international exchange of best practices and responsible AI use.
- Promote more exchanges and a cultural dialogue with tech actors to foster a mutual understanding.
- Draw insights from current practices in tech diplomacy to gain new ideas and adopt lessons learned that could be applied to work in ICR.
- Foster greater dialogue among the different AI subcultures—engineers, artists, policymakers and ethics researchers, among others.
- Explore the potential for artists and creative industries to use AI as a tool to enhance ICR.
- Shift from "it" (AI) to "we"—adopting a truly participatory, collaborative and co-construction approach.
- Collaborate on international mobility initiatives in AI fields—for researchers, artists, cultural entrepreneurs, among others—, prioritising the needs of professionals from the Global South.
- Foster not only diversity of data, content and languages but also of actors—i.e. the participation of a wide range of stakeholders, including local artists, entrepreneurs, policymakers and technologists, among others.
- Seek input from underrepresented communities.
- Encourage multidisciplinary work by internal teams across all types of organisations, including public, private and civil society.
- Encourage more frequent contacts between internal IT teams and teams developing initiatives in ICR.
- Consider the establishment of an internal advisory committee within the institution, focused on AI and digital technologies to guide strategic developments.
- Given the interministerial nature of work in ICR, coordinate more joint actions with Ministries of Culture and Foreign Affairs.
- Do not limit work to conducting activities via online streaming but also organise face-to-face events, both in countries of the Global North and the Global South.

8.3 Impact on Policies

- Ensure that tech companies or government technology agencies are not the only ones making decisions about AI research and AI policy that significantly affect ICR.
- Emphasise that the cultural dimension in this AI era should not be merely a formal addition to national AI strategies and international treaties but must be given a priority.
- Seek to strengthen the institution's participation in international debates, particularly in the UN/UNESCO, by actively engaging in discussions that directly concern cultural diversity and the protection of cultural heritage in the AI era.
- Promote international cooperation as an indispensable pillar for global AI governance.
- Ensure that the discussion on AI safety, while critical, does not overshadow other equally vital topics such as AI ethics, accountability, transparency, freedom of expression, non-discrimination, cultural diversity and fairness.
- Advocate for platforms that distribute online content to adopt a firm policy against AI-enabled harassment, malicious disinformation and other detrimental actions, as a condition for a democratic information ecosystem.
- Promote the exchange of best practices and success stories achieved in certain countries in terms of AI policy, so that other regions may also draw inspiration, while avoiding the uncritical adoption of frameworks that inherently carry local idiosyncrasies.
- Foster activities to encourage countries from the Global South to develop their own AI strategy, considering local idiosyncrasies in addition to global trends.
- Promote public policies that favour a fair and economically sustainable cultural ecosystem as a fundamental requirement for international exchange and cultural cooperation in the age of AI, including the necessity for artists to be fairly compensated for the use of their works in AI systems.
- Highlight the importance of public infrastructure initiatives in the field of AI.

8.4 New Debates

- Encourage discussions on AI ethics and governance, ensuring that principles of cultural diversity and intercultural dialogue are integral to these conversations.
- Promote critical examinations of certain trite but influential concepts, such as "superintelligence" and "rogue AI," as well as the ideologies professed by numerous tech leaders.
- Launch comprehensive studies on the imaginaries, narratives and metaphors prevalent in the AI sector, and invite all stakeholders to challenge these dominant discourses, not merely as an academic exercise but as an effective means to provoke a shift in the AI landscape.
- Explore which values, principles and action guides are indispensable when working with AI in ICR.
- Devise ways to foster a plural and democratic public debate in the digital environment, at a time when extreme polarisation, disinformation and censorship seem to impede it.
- Encourage high-level research and debates on the future impact of AI agents as well as the merger between AI and the metaverse—possible changes in concepts such as "truth" and "shared reality," among other key issues for ICR.
- Highlight the need to move beyond the narrative of AI as a race in which all countries compete not to be displaced by others and integrate other types of interaction, more rooted in cooperation, mutual understanding and learning from each other.

8.5 Impact on AI Development and Deployment

- Emphasise that diversity implies that there is no single way to develop and deploy AI systems.
- Support language-specific research communities.
- Promote transparency in AI at all levels, not only concerning data but also regarding the role of each stakeholder involved throughout the AI value chain.

- Evaluate the potential for initiating scientific research in deep learning in critical areas of ICR.
- Contribute to the creation of open datasets that are freely accessible and ethically sourced, ensuring compliance with data protection standards, while avoiding unauthorised data usage.

8.6 New Approaches

- Do not view the relationship between AI and ICR as a unidirectional phenomenon—AI impacting ICR—but rather as a multidimensional reality with reciprocal influences.
- Move away from techno deterministic perspectives exclusively centred on specific tools that seek to impose "top-down solutions," which generally do nothing more than introduce new problems.
- In any new project, promote an approach that starts from the actual needs as expressed by the local stakeholders.
- Keep the human in the loop, aware of the tools' limitations and the vested interests behind them.
- Cultivate intercultural understanding as a foundational element, recognising that a technological system cannot adequately guarantee peace, safety or fairness by itself.
- Promote new hybrid profiles, such as that of a "digital humanist," combining technology and culture.
- Avoid "reinventing the wheel"; instead, work with the most comprehensive and legitimate global frameworks (UN and UNESCO) to avoid duplication of efforts.
- Maintain agile and iterative strategies that can adapt to a perpetually changing technological context.

References

"The Bletchley Declaration by Countries Attending the AI Safety Summit" (2023). *Gov.uk*. Available at https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 [14/05/2024].

AI Forensics (2023). *Generative AI and elections: Are chatbots a reliable source of information for voters?* Available at https://aiforensics.org/uploads/Algorithm_Watch_AI_Forensics_Bing_Chat_Report_ca7200fe8d_a9bd8a6a4b.pdf [14/05/2024].

Anderson, R. (2023). "Will GPT models choke on their own exhaust?", *Light Blue Touchpaper*. Available at https://www.lightbluetouchpaper.org/2023/06/06/will-gpt-models-choke-on-their-own-exhaust [14/05/2024].

Andreesen, M. (2023). "The Techno-Optimist Manifesto", Andreesen Horowitz. Available at https://a16z.com/the-techno-optimist-manifesto [14/05/2024].

Anthropic (2023). "Claude's Constitution". Available at https://www.anthropic.com/news/claudes-constitution [14/05/2024].

Artificial Intelligence Commission (2024). *AI: Our Ambition for France.* Available at https://www.info.gouv.fr/upload/media/content/0001/09/02cbcb40c3541390be391feb3d963a4126b12598.pdf [14/05/2024].

Assayag, C. (2024). "Comment les leaders de la Silicon Valley sont devenus 'technosécessionnistes'", *Usbek & Rica*. Available at https://usbeketrica.com/fr/article/comment-les-leaders-de-la-silicon-valley-sont-devenus-technosecessionnistes [14/05/2024].

Baack, S. (2024). "Training Data for the Price of a Sandwich. Common Crawl's Impact on Generative AI", *Mozilla Insights*. Available at https://foundation.mozilla.org/documents/350/Common_Crawl_Mozilla_Foundation_2024.pdf [14/05/2024].

Bail, C.; Argyle, L.; Busby, E. (2023). "Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale", *PNAS*, vol. 120, 41. Available at https://www.pnas.org/doi/10.1073/pnas.2311627120 [14/05/2024].

Bareis, J.; Katzenbach, C. (2021). "Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics", *Science, Technology, & Human Values, vol. 47, 5*. Available at https://journals.sagepub.com/doi/10.1177/01622439211030007 [14/05/2024].

Bath, J. (2024). "Canada's Proposed Artificial Intelligence and Data Act", *Law Now*. Available at https://www.lawnow.org/canadas-proposed-artificial-intelligence-and-data-act [14/05/2024].

Bender, E.; Koller, A. (2020). "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data", *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* Available at https://aclanthology.org/2020.acl-main.463.pdf [14/05/2024].

Bender, E.; Gebru, T.; McMillan-Major, A.; Shmitchell, S. (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Available at https://dl.acm.org/doi/pdf/10.1145/3442188.3445922 [14/05/2024].

Bhuiyan, J. (2023). "Lost in AI translation: growing reliance on language apps jeopardizes some asylum applications", *The Guardian.* Available at https://www.theguardian.com/us-news/2023/sep/07/asylum-seekers-ai-translation-apps [14/05/2024].

Birhane, A. (2020). "Algorithmic Colonization of Africa", *Scripted*, vol. 17, 2. Available at https://script-ed.org/article/algorithmic-colonization-of-africa [14/05/2024].

Blume, C.; Rauchbauer, M. (2022). "How to Be a Digital Humanist in International Relations: Cultural Tech Diplomacy Challenges Silicon Valley", in Werthner, H.; Prem, E.; Lee, E. A.; Ghezzi, C. (ed.), *Perspectives on Digital Humanism,* Springer, Cham. Available at https://link.springer.com/content/pdf/10.1007/978-3-030-86144-5_15 [14/05/2024].

Bontcheva, K. (2024). *Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities*, Edmo. Available at https://veraai-cms-files.s3.eu-central-1.amazonaws.com/Generative_AI_and_Disinformation_White_Paper_v8_f784c66791.pdf [14/05/2024].

Bostrom, N. (2014). *Superintelligence: paths, dangers, strategies*, Oxford University Press, Oxford.

Bradford, A. (2023). *Digital Empires: The Global Battle to Regulate Technology,* Oxford University Press, New York.

Bremmer, I.; Suleyman, M. (2023). "The AI Power Paradox", *Foreign Affairs.* Available at https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox [14/05/2024].

Brittain, B. (2023). "AI-generated art cannot receive copyrights, US court says", *Reuters*. Available at https://www.reuters.com/legal/ai-generated-art-cannot-receive-copyrights-us-court-says-2023-08-21 [14/05/2024].

British Council; Goethe-Institut (2018). *Culture in an Age of Uncertainty: The Value of Cultural Relations in Societies in Transition*. Available at https://www.goethe.de/resources/files/pdf165/culture_in_an_age_of_uncertainty.pdf [14/05/2024].

Brodsky, S. (2024). "Can AI Be Conscious?", *AI Business*. Available at https://aibusiness.com/ml/can-ai-be-conscious- [14/05/2024].

Browne, R. (2023). "Bosch, SAP and others back Europe's answer to OpenAI with $500 million of fresh funds", *CNBC*. Available at https://www.cnbc.com/2023/11/06/aleph-alpha-raises-500-million-to-build-a-european-rival-to-openai.html [14/05/2024].

Cazzaniga, M.; Jaumotte, F.; Li, L.; Melina, G.; Panton, A. J.; Pizzinelli, C.; Rockall, E.; Tavares, M. M. (2024). "Gen-AI: Artificial Intelligence and the Future of Work", *International Monetary Fund*. Available at https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379 [14/05/2024].

Chakravorti, B. (2023). "The AI Regulation Paradox", *Foreign Policy*. Available at https://foreignpolicy.com/2023/08/04/ai-regulation-artificial-intelligence-democracy-elections [14/05/2024].

Chang, V. (2020). *Art + Tech Report 2020*, EUNIC. Available at https://09227776-4c3f-45f5-9037-85b84a0e59af.filesusr.com/ugd/f31f46_8151ab52dbe040a6af559dc574be57d3.pdf [14/05/2024].

Cho, W. (2024). "Authors Guild Exploring Blanket License for Artificial Intelligence Companies", *The Hollywood Reporter*. Available at https://www.hollywoodreporter.com/business/business-news/authors-guild-exploring-blanket-license-artificial-intelligence-companies-1235785941 [14/05/2024].

Chowdhury, R. (2024). *Technology-Facilitated Gender-Based Violence in an Era of Generative AI*, UNESCO. Available at https://unesdoc.unesco.org/ark:/48223/pf0000387483 [14/05/2024].

Christian, A. (2023). "FTX's Sam Bankman-Fried believed in 'effective altruism'. What is it?", *BBC*. Available at https://www.bbc.com/worklife/article/20231009-ftxs-sam-bankman-fried-believed-in-effective-altruism-what-is-it [14/05/2024].

Clark, K. (2023). "AI Startup Midjourney Expects $200 Million in Revenue", *The Information*. Available at https://www.theinformation.com/briefings/ai-startup-midjourney-expects-200-million-in-revenue [14/05/2024].

Climate Action (2024). *Artificial Intelligence Threats to Climate Change*. Available at https://foe.org/wp-content/uploads/2024/03/AI_Climate_Disinfo_v6_031224.pdf [14/05/2024].

CNTI (2024). *Most "Fake News" Legislation Risks Doing More Harm Than Good Amid a Record Number of Elections in 2024*. Available at https://innovating.news/wp-content/uploads/2024/04/FakeNews_Legislation_Report_Final-1.pdf [14/05/2024].

Cobb, M. (2020). "Why your brain is not a computer", *The Guardian.* Available at https://www.theguardian.com/science/2020/feb/27/why-your-brain-is-not-a-computer-neuroscience-neural-networks-consciousness [14/05/2024].

Cong, W.; Yeping, Y. (2023). "China launches Global AI Governance Initiative, offering an open approach in contrast to US blockade", *Global Times.* Available at https://www.globaltimes.cn/page/202310/1300092.shtml [14/05/2024].

Cook, S. (2023). "China's Censors Could Shape the Future of AI-Generated Content", *The Diplomat*. Available at https://thediplomat.com/2023/02/chinas-censors-could-shape-the-future-of-ai-generated-content [14/05/2024].

Crawford, K. (2021). *Atlas of AI*, Yale University Press, New Haven.

Creamer, E. (2023). "Amazon restricts authors from self-publishing more than three books a day after AI concerns", *The Guardian*. Available at https://www.theguardian.com/books/2023/sep/20/amazon-restricts-authors-from-self-publishing-more-than-three-books-a-day-after-ai-concerns [14/05/2024].

Crotti, R.; Geiger, T.; Ratcheva, V.; Zahidi, S. (2019). *Global Gender Gap Report 2020*, World Economic Forum. Available at https://www3.weforum.org/docs/WEF_GGGR_2020.pdf [14/05/2024].

Cyberspace Administration of China (2022). "Deep Synthesis Regulation". Available at https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm [14/05/2024].

Cyberspace Administration of China (2023). "Administrative Measures for Generative AI Services". Available at https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm [14/05/2024].

Dans, E. (2024). "Let's not get lost in translation: why learning a language is a good thing", *Medium*. Available at https://medium.com/enrique-dans/lets-not-get-lost-in-translation-why-learning-a-language-is-a-good-thing-c16baa672914 [14/05/2024].

Davidson, J. (2023). Google Brain founder says big tech is lying about AI extinction danger", *Financial Review.* Available at https://www.afr.com/technology/google-brain-founder-says-big-tech-is-lying-about-ai-human-extinction-danger-20231027-p5efnz [14/05/2024].

Davis, T.; Pillay, T.; Hunter, M. (2023). *AI Governance for Africa,* Thomson Reuters Foundation. Available at https://www.trust.org/dA/97390870db/pdfReport/AI%20Governance%20for%20Africa%20Toolkit%20-%20Part%201%20and%202.pdf [14/05/2024].

Dawn, O.; Bowen, B.; Camargo, A.; Richardson, M.; Robinson, J. M. (2024). "The Language Stripes", *Finding Nature.* Available at https://findingnature.org.uk/2024/01/23/the-language-stripes/#:~:text=The%20Language%20Stripes%20follow%20the,to%202080%20and%2021504 [14/05/2024].

De Vries, A. (2023), "The growing energy footprint of artificial intelligence", *Joule*, vol. 7, 10, Cell Press, Cambridge (MA).

DiBenedetto, C. (2023). "AI's exclusion of Indigenous voices is history repeating itself", *Mashable.* Available at https://mashable.com/article/ai-bias-live-transcription-indigenous-languages [14/05/2024].

Dillet, R. (2024). "Microsoft made a $16M investment in Mistral AI", *TechCrunch*. https://techcrunch.com/2024/02/27/microsoft-made-a-16-million-investment-in-mistral-ai [14/05/2024].

Doctorow, C. (2023a). "What Kind of Bubble is AI?", *Locus*. Available at https://locusmag.com/2023/12/commentary-cory-doctorow-what-kind-of-bubble-is-ai [14/05/2024].

Doctorow, C. (2023b). "The real AI fight", *Pluralistic.* Available at https://pluralistic.net/2023/11/27/10-types-of-people [14/05/2024].

Dotson, K. (2023). "Chinese AI startup Baichuan raises $300M from Alibaba, Tencent and Xiaomi", *SiliconANGLE*. Available at https://siliconangle.com/2023/10/17/chinese-ai-startup-baichuan-raises-300m-alibaba-tencent-xiaomi [14/05/2024].

Drucker, J. (2011). "Humanities Approaches to Graphical Display", *Digital Humanities Quarterly.* Available at https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html [14/05/2024].

Edwards, B. (2023). "OpenAI execs warn of 'risk of extinction' from artificial intelligence in new open letter", *Ars Technica.* Available at https://arstechnica.com/information-technology/2023/05/openai-execs-warn-of-risk-of-extinction-from-artificial-intelligence-in-new-open-letter [14/05/2024].

European Parliament (2024). "Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence". Available at https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf [14/05/2024].

Feldstein, S. (2019). "How artificial intelligence systems could threaten democracy", *The Conversation*. Available at https://theconversation.com/how-artificial-intelligence-systems-could-threaten-democracy-109698 [14/05/2024].

Floridi, L. (2023). "On the Brussels-Washington consensus about the legal definition of Artificial Intelligence", *Philos. Technol.* 36, 87, Springer. Available at https://link.springer.com/content/pdf/10.1007/s13347-023-00690-z.pdf [14/05/2024].

Floridi, L.; Nobre, A. (2024). "Anthropomorphising machines and computerising minds: the crosswiring of languages between Artificial Intelligence and Brain & Cognitive Sciences", CEDE. Available at https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4738331_code2644503.pdf?abstractid=4738331&mirid=1 [14/05/2024].

Gaida, J.; Wong-Leung, J.; Robin, S.; Cave, D. (2023). The global race for future power, *ASPI*. Available at https://ad-aspi.s3.ap-southeast-2.amazonaws.com/2023-08/ASPIs%20Critical%20Technology%20Tracker.pdf?VersionId=nVmWySgLSX2FMaS1U.uQVgQvvd_W427G [14/05/2024].

Gallardo, C. (2024). "Spain to develop open-source LLM trained in Spanish, regional languages", *Sifted*. Available at https://sifted.eu/articles/spain-large-language-model-generative-ai [14/05/2024].

Garcia, Eugenio V. (2024). "The Other in the machine: diplomacy and the AI conundrum", *Place Branding and Public Diplomacy*, Palgrave Macmillan, London.

Gebru, T. (2024). "Panel will advise Department of Homeland Security", *LinkedIn*. Available at https://www.linkedin.com/feed/update/urn:li:share:7189671982878089216 [14/05/2024].

Gebru, T.; Torres, É. (2024). "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence", *First Monday*. Available at https://firstmonday.org/ojs/index.php/fm/article/view/13636/11599 [14/05/2024].

Goldmedia (2024). *AI and Music,* SACEM and GEMA. Available at https://www.goldmedia.com/fileadmin/goldmedia/Studie/2023/GEMA-SACEM_AI-and-Music/AI_and_Music_GEMA_SACEM_Goldmedia.pdf [14/05/2024].

Google (2022). "What is Artificial Intelligence (AI)?". Available at https://cloud.google.com/learn/what-is-artificial-intelligence [14/05/2024].

Government of the People's Republic of China (2022). "Algorithmic Recommendation Regulation". Available at https://www.gov.cn/zhengce/zhengceku/2022-01/04/content_5666429.htm [14/05/2024].

Greenacre, M. (2024). "EU and US to work together on AI guidelines", *Science Business.* Available at https://sciencebusiness.net/news/ai/eu-and-us-work-together-ai-guidelines [14/05/2024].

Grover, J. (2023). "India to develop its own sovereign AI Infrastructure: Rajeev Chandrasekhar", *Financial Express.* Available at https://www.financialexpress.com/business/digital-transformation-india-to-develop-its-own-sovereign-ai-infrastructure-rajeev-chandrasekhar-3321291 [14/05/2024].

Hadgu, A. T.; Azunre, P.; Gebru, T. (2023). "Combating Harmful Hype in Natural Language Processing", ICLR. Available at https://pml4dc.github.io/iclr2023/pdf/PML4DC_ICLR2023_39.pdf [14/05/2024].

Hammer, M. (2024). "OpenAI drops ban on military tools to partner with the Pentagon", *Semafor.* Available at https://www.semafor.com/article/01/16/2024/openai-is-working-with-the-pentagon-on-cybersecurity-projects [14/05/2024].

Hao, K. (2022). "Artificial intelligence is creating a new colonial world order", *MIT Technology Review.* Available at https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism [14/05/2024].

Harris, D. E. (2024). "Open-Source AI Is Uniquely Dangerous", *IEEE Spectrum*. Available at https://spectrum.ieee.org/open-source-ai-2666932122 [14/05/2024].

Hart, R. (2023). "Forget Musk's Martian Ambition—Jeff Bezos Thinks Humans Should Live In Giant Cylindrical Space Stations", *Forbes.* Available at https://www.forbes.com/sites/roberthart/2023/12/15/forget-musks-martian-ambition-jeff-bezos-thinks-humans-should-live-in-giant-cylindrical-space-stations [14/05/2024].

Heath, R. (2023). "Democracy isn't ready for its AI test", *Axios*. Available at https://www.axios.com/2023/05/11/democracy-ai-artificial-intelligence-2024-elections [14/05/2024].

Heath, R.; Fischer, S. (2024). "FTC chair Khan: Stop monopolies before they happen", *Axios*. Available at https://www.axios.com/2024/02/13/lina-khan-monopolies-ai-media-ftc [14/05/2024].

Hochman, N; Reben, A. (2019). *Art + Tech Report 2019*, EUNIC. Available at https://www.eunicglobal.eu/media/site/24283ce432-1716830417/art-tech-report-2019_eunic-silicon-valley.pdf [14/05/2024].

Holy See Press Office (2024). "Message of the Holy Father Francis for the 58th World Day of Social Communications, 24.01.2024", *The Vatican*. Available at https://press.vatican.va/content/salastampa/en/bollettino/pubblico/2024/01/24/240124b.html [14/05/2024].

Holz, D. (2024). "We should be expecting a billion humanoid robots on Earth", *Twitter/X* (16/01/2024). Available at https://x.com/DavidSHolz/status/1747370905331015797 [14/05/2024].

Höne, K. E.; Anđelković, K.; Perućica, N.; Saveska, V.; Hibbard, L.; Maciel, M. (2019). *Mapping the challenges and opportunities of artificial intelligence for the conduct of diplomacy*, DiPLO. Available at https://www.diplomacy.edu/wp-content/uploads/2022/12/AI-diplo-report.pdf [14/05/2024].

Hoover, A. (2024). "If Taylor Swift Can't Defeat Deepfake Porn, No One Can", *Wired*. Available at https://www.wired.com/story/taylor-swift-deepfake-porn-artificial-intelligence-pushback [14/05/2024].

Hu, K.; Varghese, H. M. (2024). "Microsoft pays Inflection $650 million in licensing deal while poaching top talents, source says", *Reuters*. Available at https://finance.yahoo.com/news/microsoft-agreed-pay-inflection-650-165547758.html [14/05/2024].

Huand, K.; Varnham O'Regan, S. (2023). "Meta Splits Up Its Responsible AI Team", *The Information*. Available at https://www.theinformation.com/articles/meta-breaks-up-its-responsible-ai-team [14/05/2024].

Hui, X.; Reshef, O.; Zhou, L. (2023). "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market", *NYU Stern School of Business Research Paper Series*. Available at https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4527336_code2416677.pdf?abstractid=4527336&mirid=1 [14/05/2024].

Ittelson, P.; Rauchbauer, M. (2023). *Tech diplomacy practice in the San Francisco Bay Area*, DiPLO. Available at https://diplo-media.s3.eu-central-1.amazonaws.com/2023/05/TECH_DIPLOMACY_PRACTICE_in_Bay_area_report.pdf [14/05/2024].

Jackson, A. (2022). "Black employees make up just 7.4% of the tech workforce—these nonprofits are working to change that", *CNBC*. Available at https://www.cnbc.com/2022/02/24/jobs-for-the-future-report-highlights-need-for-tech-opportunities-for-black-talent.html [14/05/2024].

Jackson, A.; Gibbs, J.; Simmonds, K. (2024). "Supreme Court confirms AI cannot be an inventor under the Patents Act", *Womble Bond Dickinson*. Available at https://www.womblebonddickinson.com/uk/insights/articles-and-briefings/supreme-court-confirms-ai-cannot-be-inventor-under-patents-act [14/05/2024].

Jiang, B. (2023). "Meituan buys out AI start-up Light Year for US$281m as delivery services giant makes a move in ChatGPT space", *South China Morning Post*. Available at https://www.scmp.com/tech/article/3225976/meituan-buys-out-ai-start-light-year-us281m-delivery-services-giant-makes-move-chatgpt-space [14/05/2024].

Johnson, R. L.; Pistilli, G.; Menédez-González, N.; Dias Duran, L. D.; Panai, E.; Kalpokiene, J.; Bertulfo, D. J. (2022). "The Ghost in the Machine has an American accent: value conflict in GPT-3", *arXiv*. Available at https://arxiv.org/abs/2203.07785 [14/05/2024].

Jung, C.; Srinivasa Desikan, B. (2024). *Transformed by AI,* IPPR. Available at https://ippr-org.files.svdcdn.com/production/Downloads/Transformed_by_AI_March24_2024-03-27-121003_kxis.pdf [14/05/2024].

Kak, A.; West, S. M.; Whittaker, M. (2023). "Make no mistake—AI is owned by Big Tech", *MIT Technology Review.* Available at https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html [14/05/2024].

Kaplan, J.; McCandlish, S.; Henighan, T.; Brown, T. B.; Chess, B.; Child, R.; Gray, S.; Radford, A.; Wu, J.; Amodei, D. (2020). "Scaling Laws for Neural Language Models", *arXiv*. Available at https://arxiv.org/pdf/2001.08361 [14/05/2024].

Kilkenny, K. (2024). "Tyler Perry Puts $800M Studio Expansion on Hold After Seeing OpenAI's Sora", *The Hollywood Reporter.* Available at https://www.hollywoodreporter.com/business/business-news/tyler-perry-ai-alarm-1235833276 [14/05/2024].

Kinder, M. (2024). "Hollywood writers went on strike to protect their livelihoods from generative AI. Their remarkable victory matters for all workers", *Brookings.* Available at https://www.brookings.edu/articles/hollywood-writers-went-on-strike-to-protect-their-livelihoods-from-generative-ai-their-remarkable-victory-matters-for-all-workers [14/05/2024].

Knibbs, K. (2024). "Most Top News Sites Block AI Bots. Right-Wing Media Welcomes Them", *Wired*. Available at https://www.wired.com/story/most-news-sites-block-ai-bots-right-wing-media-welcomes-them [14/05/2024].

Knockel, J.; Ruan, L. (2022). *Bada Bing, Bada Boom*, Citizen Lab. Available at https://tspace.library.utoronto.ca/bitstream/1807/122063/1/Report%23156--bingsearch.pdf [14/05/2024].

Kobberod, B. (2024). "Age of AI", *Smithsonian*. Available at: https://www.si.edu/support/impact/age-of-ai [14/05/2024].

Kuczmarski, J. (2018). "Reducing gender bias in Google Translate", *Google*. Available at https://blog.google/products/translate/reducing-gender-bias-google-translate [14/05/2024].

Kulveit, J. (2023). "Snapshot of narratives and frames against regulating AI", *LessWrong*. Available at https://www.lesswrong.com/posts/PdcnEEE6sdgACDrEk/snapshot-of-narratives-and-frames-against-regulating-ai [14/05/2024].

Kwon, J.; Watson, I. (2023). "'The only thing we can't do is sign autographs': The rise of virtual K-pop bands", *CNN*. Available at https://www.cnn.com/style/kpop-virtual-bands-ai-intl-hnk/index.html [14/05/2024].

Kyoung Ko, H. (2024). "Analysis of AI regulatory frameworks in South Korea", *Asia Business Law Journal*. Available at https://law.asia/ai-regulatory-frameworks-south-korea [14/05/2024].

Le Scao, T.; Fan, A.; Akiki, C.; Pavlick, E.; Ilić, S. et al. (2022). "BLOOM: A 176B-Parameter Open-Access Multilingual Language Model", *arXiv*. Available at https://arxiv.org/pdf/2211.05100 [14/05/2024].

LeCun, Y. (2023). "Altman, Hassabis, and Amodei are the ones doing massive corporate lobbying at the moment", *Twitter/X* (29/10/2023). Available at https://x.com/ylecun/status/1718670073391378694 [14/05/2024].

Leswing, K. (2023). "Google, Amazon, Nvidia and other tech giants invest in AI startup Hugging Face, sending its valuation to $4.5 billion", *CNBC*.

Available at https://www.cnbc.com/2023/08/24/google-amazon-nvidia-amd-other-tech-giants-invest-in-hugging-face.html [14/05/2024].

Lima-Strong, C.; Oremus, W. (2024). "Meta's new AI labels won't solve the 'Taylor Swift problem'", *The Washington Post*. Available at https://www.washingtonpost.com/politics/2024/02/07/metas-new-ai-labels-wont-solve-taylor-swift-problem [14/05/2024].

Luccioni, S.; Strubell, E.; Jernite, Y. (2024). "Power Hungry Processing: Watts Driving the Cost of AI Deployment?", *arXiv*. Available at https://arxiv.org/pdf/2311.16863 [14/05/2024].

Maas, M. (2023). "AI is Like... A Literature Review of AI Metaphors and Why They Matter for Policy", *Institute for Law and AI*. Available at https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4612468_code4791224.pdf?abstractid=4612468&mirid=1 [14/05/2024].

Mabie, N. (2023). "Meet Michael Running Wolf, the man using AI to reclaim Native languages", *Missoulian.* Available at https://missoulian.com/news/local/meet-michael-running-wolf-the-man-using-ai-to-reclaim-native-languages/article_9cdd8db4-4074-5edb-9fa4-74773cdb9464.html [14/05/2024].

Macaulay, T. (2023). "Bedazzled by big tech, the UK's AI summit is overlooking big issues", *The Next Web.* Available at https://thenextweb.com/news/uk-ai-safety-summit-raises-concerns [14/05/2024].

MAGA War Room (2023). "President Trump in Iowa", *Twitter/X* (02/12/2023). Available at https://x.com/MAGAIncWarRoom/status/1731067876801474707 [14/05/2024].

Márquez Lartigue, R.(2023a). "Diplomacy 4.0: How Artificial Intelligence is Changing Diplomacy?", *Consular Diplomacy.* Available at https://www.consulardiplomacy.com/home/diplomacy-40-how-artificial-intelligence-is-changing-diplomacy [14/05/2024].

Márquez Lartigue, R. (2023b). "Diplomacy 4.0 Navigating the AI Era: Geopolitics, Diplomacy, and the Power of Artificial Intelligence", *Consular Diplomacy*. Available at https://www.consulardiplomacy.com/home/diplomacy-40-navigating-the-ai-era-geopolitics-diplomacy-and-the-power-of-artificial-intelligence [14/05/2024].

Mathews, J.; Garfinkle, A. (2024). "Inside the $1 billion love affair between Stability AI's 'complicated' founder and tech investors Coatue and Lightspeed—and how it turned bitter within months", *Fortune*. Available at https://fortune.com/2024/03/27/inside-stability-ai-emad-mostaque-bad-breakup-vc-investors-coatue-lightspeed [14/05/2024].

Matthews, D. (2023). "The $1 billion gamble to ensure AI doesn't destroy humanity", *Vox*. Available at https://www.vox.com/future-perfect/23794855/anthropic-ai-openai-claude-2 [14/05/2024].

MeitY (2023). *India AI 2023*. Available at https://www.meity.gov.in/writereaddata/files/IndiaAI-Expert-Group-Report-First-Edition.pdf [14/05/2024].

Meta (2022). "New AI Model Translates 200 Languages, Making Technology Accessible to More People". Available at https://about.fb.com/news/2022/07/new-meta-ai-model-translates-200-languages-making-technology-more-accessible [14/05/2024].

Metz, C.; Mickle, T.; Isaac, M. (2023a). "Before Altman's Ouster, OpenAI's Board Was Divided and Feuding", *The New York Times*. Available at https://www.nytimes.com/2023/11/21/technology/openai-altman-board-fight.html [14/05/2024].

Metz, C.; Weise, K.; Grant, N.; Isaac, M. (2023b). "How Elon Musk and Larry Page's AI Debate Led to OpenAI and an Industry Boom", *The New York Times*. Available at https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html [14/05/2024].

Metz, C.; Mickle, T. (2024). "OpenAI Completes Deal That Values the Company at $80 Billion", *The New York Times*. Available at https://www.nytimes.com/2024/02/16/technology/openai-artificial-intelligence-deal-valuation.html [14/05/2024].

Mhalla, A. (2023). "Technopolitique de l'IA : luttes idéologiques, tensions géopolitiques, espoirs démocratiques", *Institut Montaigne*. Available at https://www.institutmontaigne.org/expressions/technopolitique-de-lia-luttes-ideologiques-tensions-geopolitiques-espoirs-democratiques [14/05/2024].

Milligan, E. (2024). "UK Starts Drafting AI Regulations for Most Powerful Models", *Bloomberg*. Available at: https://www.bloomberg.com/news/articles/2024-04-15/uk-starts-drafting-ai-regulations-for-most-powerful-models [14/05/2024].

Mind the Bridge (2023). *European Innovation Economy in Silicon Valley*. Available at https://storage.googleapis.com/mtb-research.appspot.com/publications/2023-european-innovation-economy-in-silicon-valley/MTB-2023-european-innovation-economy-in-silicon-valley-report.pdf [14/5/2024].

Monroe, D. (2023). "Silicon Landlords: On the Narrowing of AI's Horizon", *The Nation*. Available at https://www.thenation.com/article/culture/ai-big-tech-monopoly [14/05/2024].

Morozov, E. (2013). *To Save Everything, Click Here: The Folly of Technological Solutionism*, PublicAffairs, New York.

Musk, E. (2022). "Worth reading", *Twitter/X* (02/08/2022). Available at https://x.com/elonmusk/status/1554335028313718784 [14/05/2024].

National Information Security Standardisation Technical Committee (2023). "Draft Basic Security Requirements for Generative AI Services". Available at https://www.tc260.org.cn/upload/2023-10-11/1697008495851003865.pdf [14/05/2024].

Newton-Rex, Ed (2023). "I've resigned from my role leading the Audio team at Stability AI", *Twitter/X* (15/11/2023). Available at https://x.com/ednewtonrex/status/1724902327151452486 [14/05/2024].

Nguyen, B. (2024). "Google's next AI is about to get trained by millions of Reddit posts", *Quartz*. Available at https://qz.com/google-reddit-ai-training-content-1851278623 [14/05/2024].

Nicholas, G.; Bhatia, A. (2023) *Lost in Translation: Large Language Models in Non-English Content Analysis*, CDT Research. Available at https://cdt.org/wp-content/uploads/2023/05/non-en-content-analysis-primer-051223-1203.pdf [14/05/2024].

Nooreyezdan, N. (2023). "India's religious AI chatbots are speaking in the voice of god—and condoning violence", *Rest of the World*. Available at https://restofworld.org/2023/chatgpt-religious-chatbots-india-gitagpt-krishna [14/05/2024].

O'Sullivan, D.; Gordon, A. (2023). "How Microsoft is making a mess of the news after replacing staff with AI", *CNN*. Available at https://edition.cnn.com/2023/11/02/tech/microsoft-ai-news/index.html [14/05/2024].

Ogura, J.; Liu, J. (2024). "Japan seeks investment in AI, semiconductors from American companies", *CNN*. Available at https://edition.cnn.com/2024/04/10/tech/japan-seeks-us-investment-ai-semiconductors-intl-hnk/index.html [14/05/2024].

ÓhÉigeartaigh, S.; Whittlestone, J.; Liu, Y.; Zeng, Yi; Liu, Z. (2020). "Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance", *Philosophy & Technology*, 33, Springer, Dordrecht. Available at https://link.springer.com/content/pdf/10.1007/s13347-020-00402-x.pdf [14/05/2024].

Okolo, C. (2023). "AI in the Global South: Opportunities and challenges towards more inclusive governance", *Brookings*. Available at https://www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance [14/05/2024].

Open Austria (2021). "Art + Tech Lab". Available at http://web.archive.org/web/20210423140713/https://www.open-austria.com/art-tech-lab [14/05/2024].

Open for Good Alliance (2023). "Open-Source AI Data Sharing: yes! Data Colonialism: no!", *Medium*. Available at https://medium.com/@openforgood/open-source-ai-data-sharing-yes-data-colonialism-no-3062a922de03 [14/05/2024].

OpenAI (2023). "Statement to the House of Lords", *UK Parliament*. Available at https://committees.parliament.uk/writtenevidence/126981/pdf [14/05/2024].

OpenAI (2024). "Video generation models as world simulators". Available at https://openai.com/index/video-generation-models-as-world-simulators [14/05/2024].

Oxford Insights (2023). "Government AI Readiness Index 2023". Available at https://oxfordinsights.com/ai-readiness/ai-readiness-index [14/05/2024].

Pariser, E. (2011). *The Filter Bubble: What The Internet Is Hiding From You*. Penguin Press, New York.

Peng, T. (2023). "Top 10 China's AI Stories in 2023: A Year-End Review", *Recode China AI*. Available at https://recodechinaai.substack.com/p/top-10-chinas-ai-stories-in-2023 [14/05/2024].

Perrigo, B. (2023a). "DeepMind's CEO Helped Take AI Mainstream. Now He's Urging Caution", *Time*. Available at https://time.com/6246119/demis-hassabis-deepmind-interview [14/05/2024].

Perrigo, B. (2023b). "OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic", *Time.* Available at https://time.com/6247678/openai-chatgpt-kenya-workers [14/05/2024].

Perrigo, B. (2024). "The UAE Is on a Mission to Become an AI Power", *Time.* Available at https://time.com/6958369/artificial-intelligence-united-arab-emirates [14/05/2024].

Piltch, A. (2023). "Plagiarism Engine: Google's Content-Swiping AI Could Break the Internet", *Tom's Hardware.* Available at https://www.tomshardware.com/news/google-sge-break-internet [14/05/2024].

Power, D. (2022). "Elon Musk is inspired by Iain Banks's utopian sci-fi novels – but he doesn't understand them", *The Telegraph*. Available at https://www.telegraph.co.uk/books/what-to-read/does-elon-musk-really-understand-books-claims-inspired [14/05/2024].

Public Infrastructure (2023). "Timnit Gebru Looks at Corporate AI and Sees a Lot of Bad Science". Available at https://publicinfrastructure.org/podcast/85-timnit-gebru [14/05/2024].

Ravindran, S. (2023). "AI often mangles African languages. Local scientists and volunteers are taking it back to school", *Science.* Available at https://www.science.org/content/article/ai-often-mangles-african-languages-local-scientists-and-volunteers-are-taking-it-back [14/05/2024].

Reisner, A. (2023). "These 183,000 Books Are Fueling the Biggest Fight in Publishing and Tech". *The Atlantic*. Available at https://www.theatlantic.com/technology/archive/2023/09/books3-database-generative-ai-training-copyright-infringement/675363 [14/05/2024].

Ressa, M.; Muratov, D.; et al. (2022). *10-Point Plan.* Available at https://10pointplan.org [14/05/2024].

Rescala, P.; Horta Ribeiro, M.; Hu, T.; West, R. (2024). "Can Language Models Recognize Convincing Arguments?", *arXiv*. Available at https://arxiv.org/pdf/2404.00750 [14/05/2024].

Reuters (2024). "Nvidia CEO Huang says countries must build sovereign AI infrastructure". Available at https://www.reuters.com/technology/nvidia-ceo-huang-says-countries-must-build-sovereign-ai-infrastructure-2024-02-12 [14/05/2024].

Richter, F. (2024). "The Most Spoken Languages: On the Internet and in Real Life", *Statista*. Available at https://www.statista.com/chart/26884/languages-on-the-internet [14/05/2024].

Robertson, D. (2023). "France makes a big push on open-source AI", *Digital Future Daily*. Available at https://www.politico.com/newsletters/digital-future-daily/2023/08/08/france-makes-a-big-push-on-open-source-ai-00110341 [14/05/2024].

Ryan-Mosley, T. (2023). "Catching bad content in the age of AI", *MIT Technology Review*. Available at https://www.technologyreview.com/2023/05/15/1073019/catching-bad-content-in-the-age-of-ai [14/05/2024].

Sadeghi, M.; Arvanitis, L.; Padovese, V.; Pozzi, G.; Badilini, S.; Vercellone, C.; Wang, M.; Brewster, J.; Huet, N.; Schimmel, B.; Fishman, Z.; Pfaller, L.; Adams, N. (2024). "Tracking AI-enabled Misinformation", *NewsGuard*. Available at https://www.newsguardtech.com/special-reports/ai-tracking-center [14/05/2024].

Sakai, K.; Tsuji, T.; Yasoshima, R. (2024). "Japan weighs regulating AI developers, following U.S. and EU", *Nikkei Asia*. Available at https://asia.nikkei.com/Business/Technology/Japan-weighs-regulating-AI-developers-following-U.S.-and-EU [14/05/2024].

Schmidt, E. (2022). "AI, Great Power Competition & National Security", *Daedalus*, vol. 151, 2 (pp. 288–298), MIT Press. Available at https://direct.mit.edu/daed/article-pdf/151/2/288/2060617/daed_a_01916.pdf [14/05/2024].

Schneier, B.; Sanders, N. (2023). "Build AI by the People, for the People", *Foreign Policy*. Available at https://foreignpolicy.com/2023/06/12/ai-regulation-technology-us-china-eu-governance [14/05/2024].

Schneier, B.; Farrell, H.; Sanders, N. E. (2023). "How Artificial Intelligence Can Aid Democracy", *Slate*. Available at https://slate.com/technology/2023/04/ai-public-option.html [14/05/2024].

Shah, S. (2023). "Sam Altman on OpenAI, Future Risks and Rewards, and Artificial General Intelligence", *Time*. Available at https://time.com/6344160/a-year-in-time-ceo-interview-sam-altman [14/05/2024].

Shead, S. (2017). "Google is opening a new AI research centre in China", *Business Insider*. Available at https://www.businessinsider.com/google-is-opening-a-new-ai-research-centre-in-china-2017-12 [14/05/2024].

Sheehan, M. (2023). "What the U.S. Can Learn From China About Regulating AI", *Foreign Policy*. Available at https://foreignpolicy.com/2023/09/12/ai-artificial-intelligence-regulation-law-china-us-schumer-congress [14/05/2024].

Shumailov, I.; Shumaylov, Z.; Zhao, Y.; Gal, Y.; Papernot, N.; Anderson, R. (2023). "The Curse of Recursion: Training on Generated Data Makes Models Forget", *arXiv*. Available at https://arxiv.org/pdf/2305.17493v2 [14/05/2024].

Siegmann, C.; Anderljung, M. (2022). *The Brussels Effect and Artificial Intelligence*, Centre for the Governance of AI. Available at https://cdn.governance.ai/Brussels_Effect_GovAI.pdf [14/05/2024].

Simonite, T. (2018). "AI Is the Future—But Where Are the Women?", *Wired*. Available at https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance [14/05/2024].

Spitale, G.; Biller-Adorno, N.; Germani, F. (2023). "AI model GPT-3 (dis)informs us better than humans", *Science Advances*, vol. 9, 26. Available at https://www.science.org/doi/reader/10.1126/sciadv.adh1850 [14/05/2024].

Suleyman, M. (2023). *The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma,* Crown, New York.

Sun, R. (2023). "The US has issued a 'landmark' executive order on AI", *LinkedIn.* Available at https://www.linkedin.com/posts/raymond-sun-64576a122_ai-artificialintelligence-ailaw-activity-7124877631321210880-m4Gy [14/05/2024].

Tapper, J. (2024). "Damien Hirst and Tracey Emin among thousands of British artists used to train AI software, Midjourney", *The Guardian*. Available at https://www.theguardian.com/technology/2024/jan/21/we-need-to-come-together-british-artists-team-up-to-fight-ai-image-generating-software [14/05/2024].

Taulli, T. (2024). "3 Most Common Problems with Small Language Models", *AI Business.* Available at https://aibusiness.com/nlp/3-most-common-problems-with-small-language-models [14/05/2024].

The Economist (2018). "In the struggle for AI supremacy, China will prevail". Available at https://www.economist.com/books-and-arts/2018/09/27/in-the-struggle-for-ai-supremacy-china-will-prevail [14/05/2024].

The White House (2022). "Blueprint for an AI Bill of Rights". Available at https://www.whitehouse.gov/ostp/ai-bill-of-rights [14/05/2024].

The White House (2023). "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence". Available at https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence [14/05/2024].

Thi Nguyen, C. (2024). "The Limits of Data", *Issues in Science and Technology*, vol. XL, 2. Available at https://issues.org/wp-content/uploads/2023/12/94-101-Nguyen-The-Limits-of-Data-Winter-2024.pdf [14/05/2024].

Torres, É. (2023). "The Acronym Behind Our Wildest AI Dreams and Nightmares", *Truthdig*, Available at https://www.truthdig.com/articles/the-acronym-behind-our-wildest-ai-dreams-and-nightmares [14/05/2024].

Tsanni, A. (2024). "Africa's push to regulate AI starts now", *MIT Technology Review*. Available at https://www.technologyreview.com/2024/03/15/1089844/africa-ai-artificial-intelligence-regulation-au-policy [14/05/2024].

Turk, V. (2023). "How AI reduces the world to stereotypes", *Rest of the World*. Available at https://restofworld.org/2023/ai-image-stereotypes [14/05/2024].

Tyler Millward, W. (2024). "Amazon Completes $4B Investment in AI Startup Anthropic", *CRN*. Available at https://www.crn.com/news/cloud/2024/amazon-completes-4b-anthropic-investment [14/05/2024].

UN (2023). "Members of the High-level Advisory Body on Artificial Intelligence". Available at https://www.un.org/en/ai-advisory-body/members [14/05/2024].

UN (2024). "General Assembly adopts landmark resolution on artificial intelligence". Available at https://news.un.org/en/story/2024/03/1147831 [14/05/2024].

UNECA (2022). "Africa's first AI research centre launched in Brazzaville, Congo", *Africa Renewal*. Available at https://www.un.org/africarenewal/magazine/march-2022/africas-first-ai-research-centre-launched-brazzaville-congo [14/05/2024].

UNESCO (2021). *Recommendation on the Ethics of Artificial Intelligence*. Available at https://unesdoc.unesco.org/ark:/48223/pf0000381137 [14/05/2024].

UNESCO-IRCAI (2024). *Challenging systematic prejudices: an Investigation into Gender Bias in Large Language Models.* Available at https://unesdoc.unesco.org/ark:/48223/pf0000388971 [14/05/2024].

Usbek & Rica (2023). "Musique : quand l'IA ressuscite et crée de nouveaux artistes". Available at https://usbeketrica.com/fr/article/musique-quand-l-ia-ressuscite-et-cree-de-nouveaux-artistes [14/05/2024].

Vance, A. (2024). "Sport-Coders Get Serious With Advanced AI Software Assistant", *Bloomberg*. Available at https://www.bloomberg.com/news/newsletters/2024-03-12/cognition-ai-s-devin-assistant-can-build-websites-videos-from-a-prompt [14/05/2024].

Villalobos, P.; Sevilla, J.; Heim, L.; Besiroglu, T.; Hobbhahn, M.; Ho, A. (2022). "Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning", *arXiv*. Available at https://arxiv.org/abs/2211.04325 [14/05/2024].

Vincent, J. (2022). "Google plans giant AI language model supporting world's 1,000 most spoken languages", *The Verge.* Available at https://www.theverge.com/2022/11/2/23434360/google-1000-languages-initiative-ai-llm-research-project [14/05/2024].

Vipra, J.; Korinek, A. (2023). *Market Concentration Implications of Foundation Models: The Invisible Hand of ChatGPT*, Brookings. Available at https://www.brookings.edu/wp-content/uploads/2023/09/market-concentration-implications-of-foundation-models-final-1.pdf [14/05/2024].

Walla, K. (2024). "If the US and EU don't set AI standards, China will first, say Gina Raimondo and Margrethe Vestager", *Atlantic Council.* Available at https://www.atlanticcouncil.org/blogs/new-atlanticist/if-the-us-and-eu-dont-set-ai-standards-china-will-first-say-gina-raimondo-and-margrethe-vestager [14/05/2024].

Wanat, Z. (2023). "'EU's AI act could kill our company,' says Mistral's Cédric O", *Sifted*. Available at https://sifted.eu/articles/eu-ai-act-kill-mistral-cedric-o [14/05/2024].

Weatherbed, J. (2024). "Adobe's new prototype generative AI tool is the 'Photoshop' of music-making and editing", *The Verge*. Available at: https://www.theverge.com/2024/2/28/24085551/adobe-project-music-genai-control-prototype-tool-hot-pod [14/05/2024].

Weatherby, L. (2023). "OpenAI: Metaphysics in the C-Suite", *Jacobin*. Available at https://jacobin.com/2023/11/openai-sam-altman-chatgpt-artificial-intelligence-big-tech-alignment [14/05/2024].

Weil, E. (2023). "You are not a parrot and a chatbot is not a human. And a linguist named Emily M. Bender is very worried what will happen when we forget this", *Intelligencer*. Available at https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html [14/05/2024].

Weise, K.; Metz, C. (2023). "When A.I. Chatbots Hallucinate", *The New York Times.* Available at https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html [14/05/2024].

Werthner, H. et al. (2019). *Vienna Manifesto on Digital Humanism*. Available at https://caiml.org/dighum/dighum-manifesto [14/05/2024].

West, M.; Kraut, R.; Han, C. E. (2019). *I'd blush if I could: closing gender divides in digital skills through education*, *UNESCO*. Available at https://unesdoc.unesco.org/ark:/48223/pf0000367416 [14/05/2024].

Will, J. (2024). "Could AI help save indigenous languages?", *Work in Progress.* Available at https://blog.dropbox.com/topics/work-culture/could-ai-help-save-indigenous-languages [14/05/2024].

Wintour, P. (2023). "Kamala Harris to call for urgent action on AI threat to democracy and privacy", *The Guardian.* Available at https://www.theguardian.com/technology/2023/nov/01/kamala-harris-to-call-for-urgent-action-on-ai-threat-to-democracy-and-privacy [14/05/2024].

Wirtschafter, V. (2024). "The impact of generative AI in a global election year", *Brookings*. Available at https://www.brookings.edu/articles/the-impact-of-generative-ai-in-a-global-election-year [14/05/2024].

Wodecki, B. (2024). "Geoff Hinton: Russia and US Under Trump Can Let AI 'Go Too Far'", *AI Business*. Available at https://aibusiness.com/responsible-ai/geoff-hinton-china-russia-and-us-could-let-ai-go-too-far- [14/05/2024].

Wong, M. (2024). "The AI Revolution Is Crushing Thousands of Languages", *The Atlantic*. Available at https://www.theatlantic.com/technology/archive/2024/04/generative-ai-low-resource-languages/678042 [14/05/2024].

Yagoda, M. (2024). "Airline held liable for its chatbot giving passenger bad advice – what this means for travellers", *BBC*. Available at https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know [14/05/2024].

Yong, Z.; Menghini, C.; Bach S. H. (2024). "Low-Resource Languages Jailbreak GPT-4", *arXiv*. Available at https://arxiv.org/abs/2310.02446 [14/05/2024].

Appendix: List of Interviewees

- **Aguerre, Carolina** (Uruguay). Professor at the Universidad Católica del Uruguay
- **Andrews, Hannah** (UK). Director of Digital Innovation in the Arts, British Council
- **Delpoux, Olivier** (France). Head of Digital Creation and Audiovisual Department, Institut Français
- **Garcia, Eugenio V.** (Brazil). Deputy Consul of Brazil in San Francisco
- **Jurková, Jitka** (Czech Republic). Director General, Czech Centres
- **Kastner, Stefanie** (Germany). Head of the Libraries Division, Goethe-Institut
- **Neustadt, Jeannette** (Germany). Advisor to the Secretary General, Goethe-Institut
- **Pistilli, Giada** (Italy/France). Principal Ethicist, Hugging Face
- **Prestes, Edson** (Brazil). Head of ϕ-Robotics Research Group, Federal University of Rio Grande do Sul
- **Rauchbauer, Martin** (Austria). Director at the Djerassi Resident Artists Program
- **Ruttkamp, Emma** (South Africa). Head of Philosophy at University of Pretoria and Ethics and Philosophy of AI Lead at the Centre for AI Research
- **Sekhar, Anupama** (India/UAE). Director of Policy and Engagement, International Federation of Arts Councils and Cultural Agencies (IFACCA)
- **Tomás, Isabella** (Austria). Austrian Tech Envoy, Open Austria
- **Weisinger, Maya** (US/Germany). Project Coordinator, Culture Action Europe

About the Author

Octavio Kulesz is a philosopher and a digital publisher. As a researcher collaborating with UNESCO, OIF, IFACCA and other organisations, his work focuses on digital culture. Some of his articles, such as "Culture, Platforms and Machines" (UNESCO, 2018), anticipated the contemporary challenges associated with generative AI. In 2020, he was selected by UNESCO, along with 23 other experts, to draft the Recommendation on the Ethics of AI, the world's first standard-setting instrument in this area.
Contact: ok@editorialteseo.com

Imprint

The study is created within the framework of ifa's Research Programme "Culture and Foreign Policy" and is published in the ifa Edition Culture and Foreign Policy.
The Research Programme is funded by the Federal Foreign Office.
The views expressed are those of the author and do not necessarily reflect those of ifa.

Publisher:
ifa (Institut für Auslandsbeziehungen e. V.)
Charlottenplatz 17
70173 Stuttgart
P.O. Box 10 24 63
D-70020 Stuttgart,
info@ifa.de, www.ifa.de

Authors:
Octavio Kulesz

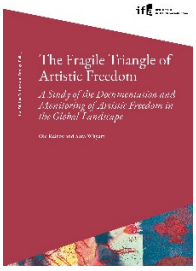
Editing:
ifa's Research Programme "Culture and Foreign Policy" and Britta Fietzke

Photo Credits:
Google DeepMind/Pexels
An artist's illustration of artificial intelligence (AI)
This image visualises the benefits and flaws of large language models. It was created by Tim West as part of the Visualising AI project.

Design:
Andreas Mayer, Stuttgart

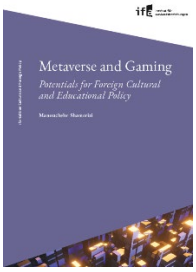
DOI: https://doi.org/10.17901/1203

Other ifa Edition Culture and Foreign Policy publications



Ole Reitov, Sara Whyatt: The Fragile Triangle of Artistic Freedom. A Study of the Documentation and Monitoring of Artistic Freedom in the Global Landscape, Stuttgart: ifa, 2024 (ifa Edition Culture and Foreign Policy). https://doi.org/10.17901/1201

Helmut K. Anheier, Edward L. Knudsen, Regina A. List: Soft Power and the New Geopolitics: Germany in Comparative Perspective, Stuttgart: ifa, 2023 (ifa External Cultural Policy Monitor). https://doi.org/10.17901/974



Manouchehr Shamsrizi: Metaverse and Gaming. Potentials for Foreign Policy and International Cultural Relations, Stuttgart: ifa, 2023 (ifa Edition Culture and Foreign Policy). https://doi.org/10.17901/957

Stefan Meister (ed.): Understanding Russian Communication Strategy. Case Studies of Serbia and Estonia, Stuttgart: ifa, 2018 (ifa Edition Culture and Foreign Policy). https://doi.org/10.17901/AKBP1.02.2018

Further publications are available at: culturalrelations.ifa.de/en/research/

Artificial Intelligence and International Cultural Relations

Challenges and Opportunities for Cross-Sector Collaboration

The integration of AI systems has profoundly transformed multiple dimensions of international cultural relations (ICR). Notably, changes in recent years have accentuated imbalances among languages, affected the situation of cultural sectors, and impacted the integrity of information while also providing a platform for influential tech players with unique imaginaries. This is unfolding in an era characterised by an increasing geopolitical competition and a global search for new AI governance frameworks. Most of these issues will likely further intensify in the coming years, with the proliferation of AI agents with whom users will interact in various contexts. In such a scenario, it has become urgent to outline a strategy for the work in the field of ICR in the AI era. Considering the mutual influence between AI and ICR, institutions engaged in ICR should aim to ensure ethical and responsible technology use and emerge as pivotal contributors to the advancement and regulation of AI.

ifa Institut für Auslandsbeziehungen