

# Voluminous yet Vacuous? Semantic Capital in an Age of Large Language Models

Luca Nannini<sup>1,2</sup>

<sup>1</sup>Minsait by Indra Sistemas SA, 35 Avenida de Bruselas, Alcobendas, Madrid, 28108, Spain

<sup>2</sup>Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) Universidade de Santiago de Compostela, Rúa de Jenaro de la Fuente Domínguez, Santiago de Compostela, 15782, Spain

## Abstract

Large Language Models (LLMs) have emerged as transformative forces in natural language processing, wielding the power to generate human-like text. However, despite their potential for content creation, they carry the risk of eroding our Semantic Capital (SC) - the collective knowledge within our digital ecosystem - thereby posing diverse social epistemic challenges. This study explores these models' evolution, capabilities, and limitations while highlighting the ethical concerns they raise. The contribution is two-fold. First, it is acknowledged that withstanding the challenges of tracking and controlling LLM impacts, we should reconsider our interaction with these AI technologies and the narratives that form their public perception. It is argued that before achieving this goal, it is essential to confront a potential epistemic tipping point in an increasing AI-driven infosphere. This goes beyond just adhering to AI ethical norms or regulations and requires understanding the spectrum of social epistemic risks LLMs might bring to our collective SC. Secondly, building on Luciano Floridi's taxonomy for SC risks, those are mapped within the functionality and constraints of LLMs. By this outlook, we aim to protect and enrich our SC while fostering a collaborative environment between humans and AI that augments and not jeopardizes human heuristics.

## Keywords

Large Language Models (LLMs), Semantic Capital (SC), Social Epistemic Risks, Human-AI Collaboration

## 1. Introduction

The fable of Funes the Memorious, conceived by Jorge Luis Borges, serves as a powerful metaphor for the era we live in. Funes, the character blessed—or rather, cursed—with perfect memory, found himself submerged in an ocean of unfiltered details. He was a prisoner of his own capacity, drowning in his universe of relentless particulars. The individual who once boasted the greatest memory lost his ability to discern the important from the trivial, transforming his mind into a “garbage heap” of excessive detail. As Borges wrote, “*To think is to forget differences, generalize, make abstractions. In the teeming world of Funes, there were only details, almost immediate in their presence*” [1]. In an echoing resonance to Funes' plight, our society now finds itself amidst a surge of information generation and consumption with uncharted challenges to our epistemic filters.

---


*Ethics and Trust in Human-AI Collaboration: Socio-Technical Approaches, August 21, 2023, MACAO, China*

✉ lnannini@minsait.com (L. Nannini)

🆔 0000-0002-4733-9760 (L. Nannini)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

This era, marked by the infinite details of our rapidly expanding infosphere, mirrors Funes' predicament. In this context, the concept of **Semantic Capital** (SC), coined by Luciano Floridi, gains paramount importance. SC can be defined as the collective information resources—knowledge, skills, or competencies—that individuals or entities possess. These resources can be harnessed to create value within our interconnected global information ecosystem [2]. This construct actively shapes the infosphere, catalyzing communication, fostering innovation, and driving informed decision-making [3].

As the realms of human cognition and artificial intelligence (AI) increasingly coalesce, their intersection is redefining the landscapes of collaboration, decision-making, and knowledge creation. In these emerging dynamics, the role of SC escalates in importance and complexity. The collaborative environments entailing human and AI integration go beyond mere task execution. They embody intricate interactions that should be dependent on mutual beneficial augmentation, and not replacement or mimicry [4, 5].

Nonetheless, this surge in information, fueled in part by AI, has the potential to generate a cascade of cognitive and sociotechnical risks e.g., cognitive overload, misinformation, social polarization, and erosion of public trust. This might lead to an epistemic 'tipping point'—an inflection where our moral obligation to promote open dissemination of AI information might conflict with our duty to prevent harm. Indeed, the relentless acceleration and proliferation of AI information might soon manifest their most detrimental and repressive effects [6, 7, 8, 9].

This paper endeavors to delve into the role of SC within the sphere of human-AI interaction. Sec.2 departs by defining its value to foster societal knowledge and trust within the context of the information ethics challenges. In Sec.3, particular attention is given to generative AI systems, particularly Large Language Models (LLMs). By contextualizing the debate happening over their capabilities and limitation, a broader range of ethical and deontological implications of LLMs are addressed in Sec.4. Central to this endeavor is a necessary reframing of AI narratives, with a mindful consideration of who benefits from these narratives and how they shape public perception. In an era where our infosphere is populated with increasingly accessible AI-generated content, a critical reassessment of our relationship with open-source practices is discussed in Sec.5. To strengthen that, it is exhorted to consider an epistemic tipping point in an AI-driven infosphere. This approach entails moving beyond calls to adhere to AI ethical guidelines or regulations and conceiving a range of social epistemic risks that LLMs pose to our collective SC. Following Floridi's taxonomy for SC risk, in Sec.6 the main contribution lies in mapping them within the capabilities and limitations of LLMs. This novel perspective moves beyond calls for AI ethics guidelines or regulations. It encourages strategies to reinforce epistemic defenses amid AI proliferation. The discourse aims to guide an equitable, sustainable infosphere where innovation and societal well-being co-exist.

## 2. Appraising Semantic Capital

Our exploration of the crucial role of SC in human-AI collaboration begins with Luciano Floridi's philosophy of information. Floridi's seminal work, birthed from the metamorphosis of the information age, places information at the core of our world understanding [10]. The *infosphere*, Floridi proposes, is an immersive information environment housing all informational

entities—humans, artificial agents, and other organisms [10]. In this sphere, constant streams and exchanges of information form a complex interaction network, shaping our reality perception and directing our actions.

Alongside this infosphere nests SC—value derived from meaningful information. It transcends mere data accumulation, presenting as well-formed, meaningful data that bolsters one’s power to create meaning—*semanticise*. An individual’s, group’s, or society’s SC stock, demonstrated in various forms like knowledge repositories, skills, shared societal norms, cultural narratives, etc., is employed and invested in information creation, understanding, and dissemination. This process fuels essential life aspects like communication, decision-making, learning, problem-solving, and others. SC’s value is intrinsically linked to its ability to enrich our understanding, navigation, and shaping of our realities. As such, managing and curating SC is vital in our increasingly information-dense society. The risks associated with SC— (a) loss, (b) unproductive, (c) underuse, or (d) misuse or (e) depreciation due to truth erosion — are defined by Floridi as “*the potential of loss of part or all of the value of some content that can no longer enhance someone’s power to semanticise something*” [10, 3].

The digital technology era has brought forth new SC dimensions. Data abundance and computational power have created pathways for enhancing and expanding our SC. AI and other digital technologies facilitate SC management and curation, aiding its effective and efficient usage and enrichment. If our world understanding is based on relationships between informational entities, and not just their intrinsic properties [11], then these technologies give rise to new SC forms that significantly impact our semanticising processes and, ultimately, shape our identities and realities<sup>1</sup>. Why, then, is it essential to highlight these concepts? Their role in shaping human-AI collaboration is central. SC provides a crucial lens through which we understand, navigate, and shape the evolving landscape of human-AI interaction in the generative AI era.

### 3. Development of LLMs and Language “Understanding”

In the compelling narrative of natural language processing (NLP), we’ve borne witness to a series of remarkable advancements over the past decade, with Large Language Models (LLMs) and other AI generative systems claiming center stage [15, 16]. Commencing with the invention of Long Short-Term Memory (LSTM) networks in 1997 [17], the journey has led to the present-day marvels of generative AI, such as GPT-4 [18].

A brief historical overview of NLP highlights the rapid progress and increasing complexity of these models. From LSTM to Word2Vec [19], from Sequence-to-Sequence models [20] to the concept of attention mechanism [21], and ultimately to the groundbreaking Transformer architecture [22] and subsequent birth of BERT [23], each evolution has refined the capacity to process and generate text, thereby influencing the constitution and use of SC.

The ‘philosophical’ foundations of these NLP applications, especially for Word2Vec, relied on

---

<sup>1</sup>SC can be differentiated from related concepts like ‘intellectual capital’ and ‘cultural capital’. While SC focuses on knowledge, skills, and resources used for communication and comprehension [3], intellectual capital pertains to an organization’s sum of knowledge and skills that provide a competitive edge [12, 13]. Cultural capital, however, refers to cultural resources like education and norms influencing individual behavior and societal opportunities [14].

concepts of Distributional Semantics [24, 25], paired with the untapped benefits and dangers of *Big Data* to mirror a presumptive realistic image of textual knowledge gathered from online repositories and communities [26, 27, 28]. Against such shallow reflection, scholars addressed concerns related to biases of this knowledge available online or within any other databases with spurious, impartial, or unguarded data entries [29, 30, 31, 32]. This raised challenges for these models, such as primarily avoid to display semantically incomplete or nonfactual information - the so-called *hallucinations* [33] in Natural Language Generation (NLG).

The advent of LLMs such as OpenAI's GPTs, and their deployment in various applications, represent the contemporary zenith of this technological trajectory [34]. These developments have profound implications for SC, raising pressing questions about the deontology of knowledge and information resources within the infosphere. Nevertheless, the rapid proliferation of these models has sparked a lively debate over their capabilities and implications.

A crucial question raised in this debate is whether LLMs genuinely understand the information processed, or if they should be considered mere *stochastic parrots*, as posited by AI researchers Emily Bender, Timnit Gebru, Angelina McMillan-Major, and (under an alias) Margaret Mitchell [35]. The paper offered a continuation of a critical inquiry toward their natural language understanding, as previously expressed in 2020 by E. Bender [36]. They argued that these models, despite their seemingly human-like text generation abilities, merely mimic patterns without comprehending the underlying meaning, potentially leading to the dilution of SC by shuffling human information pattern in a convincing manner.

Foremost, their concerns were grounded around the biases embedded in the training data, the substantial environmental footprint of training such models, and the concentration of power in a few tech giants controlling them. In echoing these concerns, Melanie Mitchell highlighted in December 2022 the limitations of LLMs in truly *understanding* the world and their reliance on superficial patterns in the data [37].

Yet, it needs to be recognized how LLMs are powerful tools that generate human-like narratives: their underlying architecture and scalability allow them to manipulate and operate inferences over the external world representations. But such abilities are generally hard to forecast, as well as to handle and interpret, by their designers. The so-called *emergent abilities*, which become more evident as the scale of the models increases, refer to the unforeseen and unplanned behavior that LLMs display, which often defy easy understanding or control by the developers themselves [38]. Such abilities can result in outputs that are surprisingly insightful or disturbingly off-mark, underscoring the unpredictability and potential risks of deploying LLMs in real-world contexts [39, 40, 41, 42]. This challenge intensifies when we consider the increasing number of studies being released for their application in practical scenarios to assist various human tasks [43, 44].

This translates to the fact that despite the property to handle a certain degree of semantic information [40] to produce coherent textual information, LLMs cannot be universally trusted as epistemic agents capable to handle pragmatic constraints of human communication. The reason for this lies in their architectures per sé, but also within potential *Eliza-effect* [37], e.g. how the user linguistically frames their prompts based on their intention and competencies [45]. This entails that the presumptive factuality of these model outputs needs then to be compared against their stochastic nature, heavily influenced by the design [46, 41] and also the interaction [47] of the users with the prompts fed.

Despite growing efforts in providing additional heuristic bases to downplay unpredictable behavior, such as with *chain-of-thought*, *constitutional AI* or *red-teaming* [38, 48, 49], a crucial question stands with the reliability and factuality of LLMs: can we equate the performances of LLMs with human understanding and knowledge? Recognizing the differences, the academic community is reevaluating how to benchmark these models' performance. This calls for more critical assessment measures that better reflect the nature and capabilities of LLMs, especially in terms of interpretability and predictability [50, 51, 46, 32].

#### 4. Fear Sells Well? On Ethical Implications of LLMs

The debate over LLMs' abilities underscores the complex implications of AI generative systems for SC. Indeed, it comes as no surprise how LLMs by their design and capabilities can profoundly influence the infosphere landscape. These models operate as powerful amplifiers and conduits of information, capable of synthesizing and generating vast amounts of text that are, in many instances, indistinguishable from human-written content.

Reasoning about their potential benefits, LLMs can democratize access to information by breaking down barriers to user understanding e.g., paraphrasing, summarizing, or translating text into different languages. By making information more accessible and interpretable, these models can enhance the inclusivity and utility of SC. Secondly, LLMs could also contribute to the expansion of SC by facilitating the creation of new content. Authors might use these tools to overcome writer's block, generate creative ideas, or automate routine writing tasks. In academic and professional settings, LLMs might help to compile emails, draft reports, write code, or even create poetry and prose, thereby enriching the diversity and volume of SC.

Such positive scenarios must be counterbalanced with a sober recognition of the potential costs and risks that these models pose to SC. Among others, the risks associated with LLMs extend beyond semantic information handling, touching upon socio-economic, political, and ethical domains, encompassing bias propagation, labor market disruptions, power centralization, misinformation campaigns, cyber threats, intellectual property issues, and unforeseen harmful uses [35, 52, 53].

At the core, LLMs can potentially spawn a proliferation of information a-like content, increasingly blurring the line with factual information. This proliferation risks diluting the quality of SC, contributing to an infosphere that is voluminous yet vacuous<sup>2</sup>.

Alongside these concerns, the susceptibility of LLMs to the propagation of false information, as explored by Bian et al., adds another layer of complexity to the debate [58]. Their study claimed how false information tends to spread and contaminate related memories in LLMs via a semantic diffusion process. Models are claimed to be subject to authority bias, often accepting false information presented in a more trustworthy style such as news or research papers. On this line, if LLMs are easily perturbable given prompt and information sources provided, they

---

<sup>2</sup>Not only related to textual abilities, but so far public opinion was surprised by the dissemination of hyperrealist portraits of public personas made through generative AI tools, e.g. Pope Francis or Donald Trump [54, 55]. Afterward, part of the public was enraged to see how a professional photographer, Boris Eldagsen, could even win an international award with an AI-produced image, or famous painters being displayed in Google's search engine alongside AI-generated imitations of their works [56, 57].

might be deployed at scale to scuffle or crowd out minor or dissenting public voices<sup>3</sup>.

Within these considerations, we should approach the paper from Bender et al. [35] as a starting point of a wider debate encompassing not only capabilities of LLMs, but rather the governance implication and social communities impacted, ultimately pertaining to the value of our shared SC [62]. Their discourse shall be not considered merely a matter of academic disquisition over semantic handling of human language. Rather, a pointed attempt to scrutinize how these generative tools are associated to a narrative about AI that serves those who possess the means and resources to develop them, capitalizing economic value and competitive advantage [12].

Through this lens, two key interpretive perspectives are discerned in this debate. The first, immediate perspective mesmerizes the public by proclaiming these LLMs as “*sparks of Artificial General Intelligence (AGI)*” [63], implying that these models display initial prototypes of human-like cognitive intelligence<sup>4</sup>. Such a view captivates public imagination and fuels, at best, a techno-optimistic narrative, while at worst, technological determinism, having public opinion feel humanity as doomed by the advent of some unavoidable, superior AI [62].

The second perspective, however, is way more sobering and less sensationalist, unpacking a far more structural and intricate argument concerning the ecology of LLMs development, commercialization, and the possession of SC in the form of know-how for gathering and maintaining increasingly sophisticated data and AI models [7]. Indeed, when the conversation frame revolves solely around the inherent risks in the models, it inadvertently diminishes the role of their developers. As Bender et al. resonated, their research served as a warning bell, cautioning against a development trajectory of AI solutions promising extraordinary capabilities without due scrutiny [35, 72].

The core issue resides in the polarization of a debate where, on one hand, one faction predominantly comprises stakeholders - such as proprietaries of AI solutions - might derive benefits from gauging public attention over these models. Their strategic maneuvers, despite genuine fears over downsides of their products, might also be geared towards maintaining the undivided attention of the global audience, intending to foster an environment conducive to the promotion and consumption of their AI-based creations. Concurrently, another group emerges posing stark opposition by unearthing the contentious aspects of such models. This group, yet widely heterogeneous, contends that these AI solutions are not inherently superior or advantageous, and instead, might cause more harm than good due to their pronounced

---

<sup>3</sup>On this note, a debate should be held on how appropriate is to deploy generative AI to represent social distress and identities, such as public manifestations [59, 60], or companies resorting to generative AI tools claiming to promote “*diversity*” through fake fashion models advertising [61], while leaning towards ethics-washing practices - failing to hire and remunerate underrepresented individuals while still leveraging their image at no costs.

<sup>4</sup>Yet, the research from Bubeck et al. is released [as for May 2023] without peer-review by a team of Microsoft and OpenAI’s researchers, using foremost controversial definitions of human intelligence as a comparison [64]. Related to the AGI narrative, Giada Pistilli, main ethicist of HuggingFace and contributor of the LLM BLOOM [65], claimed in May 2022 to not engage herself to speak any more of AGI in a fortunate Twitter thread [66]. This is because the framing of that public debate was proved only detrimental to the real harms of LLMs, cautioning an in-depth analysis of the issue in a research study published the same month [67]. This position resonates with an increasing number of scholars being cautious to adopt or even engage in using these terms in the public discourse; similarly - as also for the current paper - concerns over unnecessary anthropomorphism [68] with LLMs are now being raised while deploying terms pertaining to human cognition, such as “*hallucination*”, to address nonfactual information provided [69]. or “*dementia*” to loss of information in LLMs [70]. In this perspective, scholars are also considering making explicit design choices to prevent anthropomorphism for conversational systems [71].

socio-technical ramifications and the plausible monopoly [9] in the AI innovation landscape<sup>5</sup>.

In fact, the year 2023 witnessed an unprecedented surge in the release of LLMs applications to the public by large corporations. These developments were characterized by increasingly shortened time-to-market duration, intensifying the potential risks and implications of these systems<sup>6</sup>. This speed, while demonstrating their technological capabilities, also exposed gaps in their ethical governance. Despite their *demo* status, instances of these LLMs causing harm or harassment to users highlighted the need for careful deployment strategies and comprehensive product testing and feedback, as well as structural inquiry over the influence exerted over the AI development agenda by proprietary solutions.

Such unforeseen detrimental consequences serve as stark reminders of the need to couple AI development with comprehensive evaluation processes that prioritize societal well-being over speed and profit. Navigating this debate, one must remain cognizant of the intricate dynamics at play and question who ultimately benefits from these narratives. This to ensure that the discourse around AI and its impact on our collective SC remains grounded in empirical realities and is sensitive to the broader socio-economic implications.

## 5. Open Source and Regulations for LLMs

Let's momentarily pause and look beyond the current maelstrom of the ongoing debate on LLMs. Taking a step back, we find ourselves in the birth of the internet era, deeply influenced by the late 20th century's internet narratives. This was a time ripe with the promise of an information revolution, catalyzed by the birth of the open-source paradigm [85]. By providing a universal platform accessible to anyone with an internet connection, it was an embodiment of the democratic ethos of these emerging digital utopias.

---

<sup>5</sup>Interestingly, the fervor and dynamism of this debate have garnered widespread attention. With the current momentum, an increasing number of scholars and civil rights associations are echoing the apprehensions about the potential LLMs can inflict, taking actions such as open letters to regulate LLMs. Of this group, a segment of the public is lending credence to the "*longtermism*" outlook—holding onto the belief that AI might be a blessing for all humanity in the future, only if it is perceived as an existential threat today [73, 74]. This viewpoint, however, does not advocate for immediate and tangible action against present structural issues, such as the exploitation of underrepresented communities involved in annotating and moderating LLMs. In response to these systemic issues, the affected communities have begun showcasing innovative grassroots initiatives. Karen Hao's investigation into AI colonialism and the protests staged by African AI workers to unionize in Nairobi illuminate these ongoing efforts [75, 76]. Meanwhile, it is noteworthy that AI pioneers like Geoffrey Hinton have been vocal about the necessity for increased regulations but have not explicitly extended support to these communities or other concerned academics, such as Bender, Gebru, and Mitchell [77]. Similarly, owners of AI technologies, like OpenAI's CEO Samuel Altman, have sought regulatory measures before the US Senate [78], while other industry leaders, such as Microsoft Chief Economist Michael Schwarz [79] and former Google CEO Eric Schmidt [80], have either invited caution over the perceived risks of generative AI until incidents of "*meaningful harm*" occur or advocated for self-regulation in the industry while criticizing governments for their alleged lack of expertise to regulate technology effectively. The narrative spun by these AI proprietors oscillates between demanding no regulation and advocating for a different regulation. Such a seemingly contradictory stance might be interpreted as a strategic maneuver to hold investor attention captive while cleverly deflecting competitive threats in the AI arena [81].

<sup>6</sup>The rush to launch these applications often eclipsed necessary precautions, resulting in technology releases without sufficient safeguards. This haste raises concerns about corporate decision-making and leaves the public exposed to unanticipated AI-related risks, such as LLMs chat-bots harassing or recommending users to self-harm or indulge minors into socially irresponsible behaviors [82, 83, 84].

The open-source movement, anchored in collaboration, transparency, and accessibility, has spurred an incredible acceleration in technological evolution [86]. This movement's transformative impact is especially palpable in the AI field, cultivating a fertile ecosystem ripe for progress and innovation. Emerging in this backdrop, LLMs owe much of their rapid development to open-source AI frameworks like *TensorFlow* and *PyTorch* as well as the transformer architecture [87, 88, 22]. Such open-source tools have made it feasible for researchers, developers, and organizations across the globe to access, modify, and contribute to a shared body of knowledge and codebase.

This democratization of AI technologies, however, is a double-edged sword - while it empowers innovation and progress, it simultaneously amplifies challenges related to misuse, ethical implications, and regulatory requirements. The diffusion of generative AI technologies, such as LLMs, via open-source platforms, accentuates the dual-use risk. LLMs can be applied for both beneficial and harmful purposes. Still cognizant of their risks, once an AI model is made openly available, specularly becomes harder to track, contain, or retract, given the scale, speed, and accessibility facilitated by open-source platforms. If instead an LLM is proprietary, such as GPT-4 [18], being undisclosed to the public, then risks might arise in not being able to reprove its design phase and data provenance, as well as oversight its deployment.

From this, it comes as no surprise that regulating generative AI technologies is a formidable challenge [89, 90]. The pace at which AI evolves is often unmatched by the rate at which traditional regulatory frameworks adapt<sup>7</sup>. Crafting effective regulations requires a delicate balancing act: on one side, for disclosed models, it entails to manage the risks of misuse while preserving the democratic ethos of open-source, without stifling innovation; on the other, for proprietary models, it entails preserving marketing advantages while still imparting auditing measures to reprove model compliance and benevolence within regulatory standards.

One potential pathway forward involves revisiting our relationship with open-source practices in the context of LLMs. Strategies could include more accountable deployers' practices, having them bear a greater responsibility for their creations, and revised legal frameworks that adapt to the specific challenges of LLMs. In terms of soft-power, this could be complemented by industry-wide certifications and licensing<sup>8</sup> to enhance accountability over the design and development of those AI systems [94, 89].

In terms of hard-power, instead, AI governance measures should attain from clear legislative guardrails, such as regulatory sandboxes, risk assessments, and auditing practicing encompassing the development and deployment of LLMs [89, 90]. Within this scope, the current major regulatory effort in the global landscape is now being lead by the European Union (EU), yet not being exempted from potential legislative weaknesses that might not always efficiently mitigate LLMs risks<sup>9</sup>.

---

<sup>7</sup>An example of this challenge can be found in the EU commissions efforts back in April 2023 to make amendments targeting generative AI, ahead of final parliamentary vote on May 11th with the EU AI Act draft [91, 92].

<sup>8</sup>For licenses, a leading example is RAILS. The *BigScience* project, an open collaborative initiative, introduces a Responsible AI License (RAIL) for the usage of their LLMs to balance accessibility and risk mitigation. It reflects a community-led approach to restrict potential LLM harms, such also concerns about their societal and environmental impacts [93].

<sup>9</sup>In particular, the current amendment draft of the EU AI Act voted on May 2023 introduced definition and provisions targeting LLMs, intended as *foundation models* [95]. At the current stage of draft, Art.28b(4), although partially beneficial with its transparency obligations, is criticized for its lack of duties imposed on online AI content



## 6. Navigating the Information Surge

Such accessibility to generate information blurs the lines between reality and artificial constructs, echoing Baudrillard’s notion of “*hyperreality*”. The hyperreality conceived by Baudrillard—an environment where simulacra blur the boundaries between real and artificial, and virtual identities supersede from a deontological lens their real references—becomes an eerily accurate premonition of a possible AI-saturated infosphere [96]. Despite being awash with information, we are precariously perched on the edge of what James Bridle refers to as a “*New Dark Age*,” a paradox where our technological ecosystem obscures knowledge instead of revealing it [8].

To resist unchallenged acceptance of an AI-driven information ecosystem, calling for ethic and regulations might not be enough to shelter our epistemic filters. While this surge of information has democratized access to knowledge and fueled progress in myriad fields, it also has the potential to create a state of social epistemic bewilderment.

It is against this backdrop that it can be argued that we have reached an epistemic *tipping point*—an inflection where the relentless acceleration and proliferation of information, aided by generative AI systems, culminate in the epistemic condition to scale up its detrimental effects. Such concept suggests a juncture where our moral obligation to assist to the open dissemination of certain AI narratives and solutions may come into conflict with our duty to prevent SC risks. This tipping point is precipitated by the realization that unfettered access to these generative models can also amplify risks given their scalability and integration, independently of liability of major AI proprietors or individual developers and deployers. As AI-generated content swells, we confront the dual challenge of strengthening our cognitive ecology to preserve our SC, whilst upholding the open-source principles that have traditionally sparked innovation [97].

Such challenge aligns with Floridi’s information ethics, which underscores the moral implications of creating, managing, and utilizing information. As remarked before, Floridi stresses how that the quality of our infosphere, or the environment in which information is created, shared, and consumed, profoundly impacts our lives and our moral decisions [2, 11].

To navigate this new complex infosphere, we must engage with a multi-faceted strategy. First, it necessitates moving forward from merely calling AI systems to adhere to ethical guidelines or exhorting to establish a culture of accountability, transparency, and shared responsibility when AI proprietors are able to influence AI agenda and public opinion [9]. This shift in approach should involve a critical reexamination of why, in such informational ecology, certain narratives tends to dominate public attention, and who benefits from this status quo [98]. Such societal introspection might prompt a critical reconsideration of the merits of confining the AI debate and our notion of innovation to a single range of solutions. Furthermore, it can be argued that, while public online information sources have proven to be fertile ground for the proliferation of AI technologies, today the wealth of SC at stake might be threatened by a range of epistemic risks outlined as following using Floridi’s taxonomy [3]:

- **Loss of SC:** This occurs when there is an oversimplification of complex semantic ideas or when an LLM relies on biased or erroneous explanatory models based on incomplete

---

generators, necessary for curbing misinformation. Yet, the Act is not yet enforced, and will likely have to interplay, within the EU regulatory ecosystem, with other regulations being discussed or already enacted. For an in-depth overview of these legislative implications, also outstanding the EU ground, refer to [90].

or distorted input data, resulting in flawed argumentation [35, 58]. In this case, the value of the semantic content is reduced due to the propagation of inaccurate or misleading information, akin to the spread of propaganda, fake news, or “*alternative facts*” [52, 53]. Protection against this type of risk necessitates deployment of external knowledge bases, rigorous data curation (e.g., data provenance and lineage) and model validation protocols to ensure LLMs generate accurate and reliable information.

- **Unproductiveness and Underuse:** When LLMs are used to replicate semantic content without adding value or facilitating a deeper understanding, it can lead to the stagnation of SC. This can happen when users rely too heavily on LLMs for information generation and consumption while neglecting to actively participate in knowledge sharing and debate. Also, at the core, this underuse of SC might stem from the LLMs’ architecture, being able to fetch only data that might be available in accessible online repositories, without yet considering the ‘*long-tail*’ of secondary, related contributions, as well as different perspectives, on a given topic. To guard against this risk, it’s essential first to inquire over the role of LLMs as epistemic agents, as well as to foster a culture of critical thinking and active engagement in the discourse, preventing the ‘mummification’ of SC [97].
- **Misuse:** LLMs, if not properly calibrated or deployed by malicious actors, can generate content that disrespects, misunderstands, or illegitimately appropriates information [38, 58]. This misuse, or information expropriation, leads to the loss of SC while also reinforcing adversarial narratives. Mitigating this risk requires careful design and account over their deployment [94], with due respect for cultural nuances and contexts. In terms of data, this might be possible also leveraging underrepresented communities to not just moderate, but actively participate in data annotation policies, to mitigate potential biases [32]. In terms of models, intellectual property, trademarks, and measure to ensure accountability shall be established to track responsibility for the development and deployment of generative solutions [94, 90], also enforced by hard laws, such as the forthcoming EU AI Act or the Liability Directive [99, 95].
- **Depreciation:** The value of SC could depreciate over time, particularly when new LLM-generated information floods the infosphere and obscures or distorts earlier knowledge. Future LLMs models, being trained or fine-tuned in such a stagnating environment, might see an increase in diminished returns over their performance. This could happen by being fed data that are either synthetically produced or, even worse, being produced by a shrunken online community of users that lacks incentives to share and engage in knowledge creation and maintenance given the information accessibility of LLMs. Also connected to underuse, the concept of “*model dementia*” has been recently coined [70] to signal how future LLMs training datasets might lead to diminished returns in terms of content richness i.e., forgetting underlying data distributions.

Building on this array of risks, our collective reliance on language models as repositories of information might entail a shift in our ethical responsibilities, as we transfer the locus of our communal knowledge from the outward sphere of human discourse to the inward representations in these models.

This shift of direction needs also to be put in context with two additional factors being inversely proportional, such as availability of information and attention. With the sheer amount

of data being produced by LLMs, we might approach new states of information magnitude. This overabundance of information is overshadowing and possibly distorting pre-existing knowledge, causing the depreciation of SC. From this, it might become progressively more demanding to discern useful information or valuable knowledge in the face of this onslaught, which in turn undermines the value derived from it.

In this era of *Attention Economy* [100], where human attention is a scarce and coveted resource, the pressure on LLMs to be deployed within work or educational tasks, outreach various audiences, and produce engaging content can inadvertently contribute to this range of risks. As these models strive to produce information that appears coherent and well-expounded - such as also sensationalist AI-generated images or news of public personas, sociopolitical facts etc - the focus might shift from providing comprehensive and nuanced insights to offering quick, often shallow pieces of information. This shift could potentially “flatten” the richness of discourse, leading to apparently more engaging, yet less insightful information being circulated.

At the core of this acceleration, *epistemic filters* becomes paramount [101]. These are mechanisms that people use to sort and interpret the information they encounter. They help us decide what counts as evidence for forming a belief or what challenges it enough to lead to belief revision. Yet robust filters also underpin our collective epistemic resilience - the ability to appropriately update beliefs based on evidence. This entails maximizing our epistemic fitness - skillfully navigating new claims and ideas to reach accurate understanding [101].

From there, future conversations should tackle how LLMs could be deployed to reinforce existing viewpoints and ethical values, possibly underpinning the deployment of epistemic filters if online users will be led to believe that AI-generated information is actually factual and representative of an allegedly major group of people than it is in reality [52, 98].

The call to action is thus twofold. On one hand, consumers of AI-generated content need to refine their individual epistemic filters to navigate this new information landscape effectively. This might entail questioning why certain narratives are spread and validated, and for which purposes. On the other, developers and proprietors of LLM carry the ethical responsibility to design systems that support, rather than undermine, the collective epistemic filters. Deployers, similarly, shall use these tools cognizant of the value of public SC, being subject to watermarks, licensing, and any other enforcement to reprove accountability.

In this vein, a collective response to these risks is the amplification of AI literacy initiatives. Creating an informed citizenry that understands AI technologies, including their potential advantages and associated risks, enables individuals to engage in meaningful discussions and decision-making processes concerning their epistemic validity. Central to this endeavor is the proactive integration of ethical considerations. Ethical responsibility should not be a reactionary measure or an isolated response to negative outcomes (e.g. regulate only when meaningful harm occurs). Instead, it needs to be woven into the fabric of the AI design and deployment process. Such proactive responsibility can serve as a safeguard, aligning the development and utilization of AI technologies, and not incentivize diminishing time-to-market agendas. However, this inquiry does not suggest a departure from open-source practices. Rather, it signals the need for a matured, conscientious version of open-source, devoid of narratives and utopias of technological emancipation or determination. One that is sober, cognizant of the social epistemic risks, and dedicated to enhancing public comprehension of AI technologies.

## 7. Conclusion

This work attempts to evaluate the complex interplay between LLMs' potential for knowledge democratization and the sociotechnical challenges they present. Amid the accelerating proliferation of LLMs in 2023, the widespread narrative that frames them as precursors to AGI risks overshadowing important socio-economic implications, potentially facilitating an AI monopoly. Despite acknowledging the lively nature of this debate, this attempt explores the delicate balance between the democratization of knowledge and the emergence of an epistemic tipping point in our infosphere.

This dynamic is exacerbated by the cognitive deluge driven by AI technologies, especially LLMs, leading to uncharted social epistemic challenges that stem from their ability to craft at scale semantic knowledge. It was highlighted that the unchecked expansion and proliferation of AI-generated content such as textual information from LLMs, while holding considerable promise, also pose significant risks. Aside from the engaging debate over their properties to handle semantic information, one shall not fail to commit to a broader inquiry over the ecosystem that fuels attention towards them, being cognizant of a different array of risks that ultimately affect the value of our SC.

## Acknowledgments

Funding contribution from the ITN project NL4XAI (*Natural Language for Explainable AI*). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860621. This document reflects the views of the author(s) and does not necessarily reflect the views or policy of the European Commission. The REA cannot be held responsible for any use that may be made of the information this document contains.

A special thanks to Pietro Belloni, Ph.D., and the doctoral researchers at the Department of Statistical Sciences, University of Padova (UniPd), Italy.

## References

- [1] J. L. Borges, *Funes, the Memorious*, Duke University Press, New York, USA, 2002, pp. 306–312. doi:doi:10.1515/9780822384182-045.
- [2] L. Floridi, *Information: A Very Short Introduction*, New York: Oxford University Press, 2010.
- [3] L. Floridi, *Semantic capital: its nature, value, and curation*, *Springer Philosophy & Technology* 31 (2018) 481–497. doi:10.1007/s13347-018-0335-1.
- [4] E. Brynjolfsson, *The turing trap: The promise & peril of human-like artificial intelligence*, *CoRR abs/2201.04200* (2022). arXiv:2201.04200.
- [5] S. Vallor, *The AI mirror: Reclaiming our humanity in an age of machine thinking*, in: V. Conitzer, J. Tasioulas, M. Scheutz, R. Calo, M. Mara, A. Zimmermann (Eds.), *AIES '22: AAAI/ACM Conference on AI, Ethics, and Society*, Oxford, United Kingdom, May 19 - 21, 2021, ACM, 2022, p. 6. doi:10.1145/3514094.3539567.

- [6] B.-C. Han, *In the swarm: digital prospects*, volume 3, MIT press, 2017.
- [7] K. Crawford, *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*, Yale University Press, 2021.
- [8] J. Bridle, *New dark age: Technology and the end of the future*, Verso Books, 2018.
- [9] D. McQuillan, *Resisting AI: an anti-fascist approach to artificial intelligence*, Bristol University Press, 2022.
- [10] L. Floridi, *The Philosophy of Information*, New York: Oxford University Press, 2011. doi:10.5840/tpm20105048.
- [11] L. Floridi, *The Routledge handbook of philosophy of information*, Routledge London, 2016.
- [12] T. Stewart, N. Brealey, *Intellectual capital: The new wealth of organizations*, *Long Range Planning* 30 (1997) 953. doi:[https://doi.org/10.1016/S0024-6301\(97\)80956-9](https://doi.org/10.1016/S0024-6301(97)80956-9).
- [13] S. Firer, S. M. Williams, *Intellectual capital and traditional measures of corporate performance*, *Journal of intellectual capital* 4 (2003) 348–360. doi:10.1108/14691930310487806.
- [14] P. Bourdieu, *The forms of capital*, in: *The sociology of economic life*, Routledge, 2018, pp. 78–92.
- [15] U. Naseem, I. Razzak, S. K. Khan, M. Prasad, *A comprehensive survey on word representation models: From classical to state-of-the-art word representation language models*, *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 20 (2021). doi:10.1145/3434237.
- [16] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J. Nie, J. Wen, *A survey of large language models*, *CoRR abs/2303.18223* (2023). doi:10.48550/arXiv.2303.18223. arXiv:2303.18223.
- [17] S. Hochreiter, J. Schmidhuber, *Long short-term memory*, *Neural Comput.* 9 (1997) 1735–1780. doi:10.1162/neco.1997.9.8.1735.
- [18] OpenAI, *GPT-4 technical report*, *CoRR abs/2303.08774* (2023). doi:10.48550/arXiv.2303.08774. arXiv:2303.08774.
- [19] T. Mikolov, K. Chen, G. Corrado, J. Dean, *Efficient estimation of word representations in vector space*, in: Y. Bengio, Y. LeCun (Eds.), *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, 2013. URL: <http://arxiv.org/abs/1301.3781>.
- [20] I. Sutskever, O. Vinyals, Q. V. Le, *Sequence to sequence learning with neural networks*, in: Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, 2014*, pp. 3104–3112. URL: <https://proceedings.neurips.cc/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html>.
- [21] D. Bahdanau, K. Cho, Y. Bengio, *Neural machine translation by jointly learning to align and translate*, in: Y. Bengio, Y. LeCun (Eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL: <http://arxiv.org/abs/1409.0473>.
- [22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, *Attention is all you need*, in: I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Pro-*

- cessing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, 2017, pp. 5998–6008. URL: <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>.
- [23] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, *CoRR* abs/1810.04805 (2018). arXiv:1810.04805.
- [24] J. Firth, A synopsis of linguistic theory, 1930-1955, *Studies in linguistic analysis* (1957) 10–32.
- [25] M. Brunila, J. LaViolette, What company do words keep? revisiting the distributional semantics of J.R. firth & zellig harris, in: M. Carpuat, M. de Marneffe, I. V. M. Ruíz (Eds.), *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022, Association for Computational Linguistics, 2022*, pp. 4403–4417. doi:10.18653/v1/2022.naacl-main.327.
- [26] V. Mayer-Schönberger, K. Cukier, *Big data: A revolution that will transform how we live, work, and think*, Houghton Mifflin Harcourt, 2013.
- [27] A. Gandomi, M. Haider, Beyond the hype: Big data concepts, methods, and analytics, *Int. J. Inf. Manag.* 35 (2015) 137–144. doi:10.1016/j.ijinfomgt.2014.10.007.
- [28] M. Zook, S. Barocas, danah boyd, K. Crawford, E. Keller, S. P. Gangadharan, A. Goodman, R. Hollander, B. König, J. Metcalf, A. Narayanan, A. Nelson, F. Pasquale, Ten simple rules for responsible big data research, *PLoS Comput. Biol.* 13 (2017). doi:10.1371/journal.pcbi.1005399.
- [29] d. boyd, K. Crawford, Critical questions for big data, *Information, Communication & Society* 15 (2012) 662–679. doi:10.1080/1369118X.2012.678878.
- [30] C. S. Calude, G. Longo, The deluge of spurious correlations in big data, *Foundations of science* 22 (2017) 595–612. doi:10.1007/s10699-016-9489-4.
- [31] B. D. Mittelstadt, L. Floridi, The ethics of big data: Current and foreseeable issues in biomedical contexts, *The Ethics of Biomedical Big Data* (2016) 445–480. doi:10.1007/978-3-319-33525-4\_19.
- [32] R. Navigli, S. Conia, B. Ross, Biases in large language models: Origins, inventory and discussion, *J. Data and Information Quality* (2023). doi:10.1145/3597307, just Accepted.
- [33] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, P. Fung, Survey of hallucination in natural language generation, *ACM Comput. Surv.* 55 (2023). doi:10.1145/3571730.
- [34] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, et al., *Improving language understanding by generative pre-training*, 2018. URL: <https://openai.com/research/language-unsupervised>.
- [35] E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big?, in: M. C. Elish, W. Isaac, R. S. Zemel (Eds.), *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021, ACM, 2021*, pp. 610–623. doi:10.1145/3442188.3445922.
- [36] E. M. Bender, A. Koller, Climbing towards NLU: on meaning, form, and understanding in the age of data, in: D. Jurafsky, J. Chai, N. Schluter, J. R. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020*,

- Online, July 5-10, 2020, Association for Computational Linguistics, 2020, pp. 5185–5198. doi:10.18653/v1/2020.acl-main.463.
- [37] M. Mitchell, D. C. Krakauer, The debate over understanding in ai’s large language models, CoRR abs/2210.13966 (2022). doi:10.48550/arXiv.2210.13966. arXiv:2210.13966.
- [38] S. R. Bowman, Eight things to know about large language models, CoRR abs/2304.00612 (2023). doi:10.48550/arXiv.2304.00612. arXiv:2304.00612.
- [39] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, . S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, CoRR abs/2005.14165 (2020). arXiv:2005.14165.
- [40] B. Z. Li, M. Nye, J. Andreas, Implicit representations of meaning in neural language models, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 1813–1827. doi:10.18653/v1/2021.acl-long.143.
- [41] R. Schaeffer, B. Miranda, S. Koyejo, Are emergent abilities of large language models a mirage?, 2023. arXiv:2304.15004.
- [42] A. V. Miceli-Barone, F. Barez, I. Konstas, S. B. Cohen, The larger they are, the harder they fail: Language models do not recognize identifier swaps in python, 2023. arXiv:2305.15507.
- [43] Z. Wang, S. Cai, A. Liu, X. Ma, Y. Liang, Describe, explain, plan and select: Interactive planning with large language models enables open-world multi-task agents, CoRR abs/2302.01560 (2023). doi:10.48550/arXiv.2302.01560. arXiv:2302.01560.
- [44] Y. Xie, C. Yu, T. Zhu, J. Bai, Z. Gong, H. Soh, Translating natural language to planning goals with large-language models, CoRR abs/2302.05128 (2023). doi:10.48550/arXiv.2302.05128. arXiv:2302.05128.
- [45] E. Perez, S. Ringer, K. Lukosiute, . E. Hubinger, N. Schiefer, J. Kaplan, Discovering language model behaviors with model-written evaluations, CoRR abs/2212.09251 (2022). doi:10.48550/arXiv.2212.09251. arXiv:2212.09251.
- [46] M. Turpin, J. Michael, E. Perez, S. R. Bowman, Language models don’t always say what they think: Unfaithful explanations in chain-of-thought prompting, 2023. arXiv:2305.04388.
- [47] M. Zhang, O. Press, W. Merrill, A. Liu, N. A. Smith, How language model hallucinations can snowball, 2023. arXiv:2305.13534.
- [48] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, D. Zhou, Chain-of-thought prompting elicits reasoning in large language models, 2023. arXiv:2201.11903.
- [49] B. Wang, S. Min, X. Deng, J. Shen, Y. Wu, L. Zettlemoyer, H. Sun, Towards understanding chain-of-thought prompting: An empirical study of what matters, CoRR abs/2212.10001 (2022). doi:10.48550/arXiv.2212.10001. arXiv:2212.10001.
- [50] S. Bowman, The dangers of underclaiming: Reasons for caution when reporting how NLP systems fail, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 7484–7499. doi:10.18653/v1/2022.acl-long.516.
- [51] S. Tedeschi, J. Bos, T. Declerck, J. Hajic, D. Hershcovich, E. H. Hovy, A. Koller, S. Krek, S. Schockaert, R. Sennrich, E. Shutova, R. Navigli, What’s the meaning of superhuman

- performance in today's nlu?, 2023. arXiv:2305.08414.
- [52] R. L. Johnson, G. Pistilli, N. Menéndez-González, L. D. D. Duran, E. Panai, J. Kalpokiene, D. J. Bertulfo, The ghost in the machine has an american accent: value conflict in GPT-3, CoRR abs/2203.07785 (2022). arXiv:2203.07785.
- [53] L. Weidinger, J. Uesato, M. Rauh, S. Griffin, Conor [...] Legassick, G. Irving, I. Gabriel, Taxonomy of risks posed by language models, in: 2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 214–229. doi:10.1145/3531146.3533088.
- [54] K. Huang, Why pope francis is the star of a.i.-generated photos, 2023. URL: <https://www.nytimes.com/2023/04/08/technology/ai-photos-pope-francis.html>.
- [55] M. Novak, Viral images of donald trump getting arrested are totally fake (for now), 2023. URL: <https://www.forbes.com/sites/mattnovak/2023/03/19/viral-images-of-donald-trump-getting-arrested-are-totally-fake>.
- [56] J. Grierson, Photographer admits prize winning image was ai generated, 2023. URL: <https://www.theguardian.com/technology/2023/apr/17/photographer-admits-prize-winning-image-was-ai-generated>.
- [57] M. Harrison, Top google result for “edward hopper” an ai-generated fake, 2023. URL: <https://futurism.com/top-google-result-edward-hopper-ai-generated-fake>.
- [58] N. Bian, P. Liu, X. Han, H. Lin, Y. Lu, B. He, L. Sun, A drop of ink may make a million think: The spread of false information in large language models, 2023. arXiv:2305.04812.
- [59] L. Taylor, Amnesty international criticised for using ai-generated images, 2023. URL: <https://www.theguardian.com/world/2023/may/02/amnesty-international-ai-generated-images-criticism>.
- [60] D. Samed, Adobe stock is flooded with ai generated gay pride content, 2023. URL: <https://twitter.com/DeanSamed/status/1658833605882265602>.
- [61] M. Clark, Levi's addresses backlash after using ai models to 'increase diversity' in online shopping, 2023. URL: <https://www.independent.co.uk/life-style/fashion/levis-ai-models-diversity-backlash-b2310280.html>.
- [62] R. Boyd, R. J. Holton, Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation?, *Journal of Sociology* 54 (2018) 331–345. doi:10.1177/1440783317726591.
- [63] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, Y. Zhang, Sparks of artificial general intelligence: Early experiments with gpt-4, 2023. arXiv:2303.12712.
- [64] C. Metz, Microsoft says new a.i. shows signs of human reasoning, 2023. URL: <https://www.nytimes.com/2023/05/16/technology/microsoft-ai-human-reasoning.html>.
- [65] BigScience, T. L. Scao, A. Fan, C. Akiki, E. Pavlick, S. Ilić, [...], Y. Belkada, T. Wolf, Bloom: A 176b-parameter open-access multilingual language model, 2023. arXiv:2211.05100.
- [66] G. Pistilli, Engaging in philosophical discussions about conscious ai/superintelligent machines, 2022. URL: <https://twitter.com/GiadaPistilli/status/1530136739959951361>.
- [67] G. Pistilli, What lies behind agi: Ethical concerns related to llms, *Éthique Et Numérique* 1 (2022) 59–68.
- [68] M. Shanahan, Talking about large language models, 2023. arXiv:2212.03551.
- [69] N. Klein, Ai machines aren't “hallucinating”. but their makers are., 2023. URL: <https://www>.



- theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein.
- [70] I. Shumailov, Z. Shumaylov, Y. Zhao, Y. Gal, N. Papernot, R. Anderson, Model dementia: Generated data makes models forget, 2023. arXiv:4921052.
  - [71] G. Abercrombie, A. C. Curry, T. Dinkar, Z. Talat, Mirages: On anthropomorphism in dialogue systems, 2023. arXiv:2305.09800.
  - [72] E. Morozov, To save everything, click here: The folly of technological solutionism, Public Affairs, 2013.
  - [73] L. Eliot, Ai ethics and ai law wrestling with ai longtermism versus the here and now of ai, 2022. URL: <https://www.forbes.com/sites/lanceeliot/2022/10/25/ai-ethics-and-ai-law-wrestling-with-ai-longtermism-versus-the-here-and-now/?sh=3a4b515e1c0f>.
  - [74] P. Olson, Ai longtermism alarmists are dragging us all down existential rabbit hole, 2023. URL: <https://www.bloomberg.com/opinion/articles/2023-05-19/ai-longtermism-alarmists-are-dragging-us-all-down-existential-rabbit-hole>.
  - [75] K. Hao, An mit technology review series: Ai colonialism, 2022. URL: <https://www.technologyreview.com/supertopic/ai-colonialism-supertopic/>.
  - [76] B. Perrigo, 150 african workers for ai companies vote to unionize, 2023. URL: <https://time.com/6275995/chatgpt-facebook-african-workers-union/>.
  - [77] J. Taylor, A. Hern, 'godfather of ai' Geoffrey Hinton quits google and warns over dangers of misinformation, 2023. URL: <https://www.theguardian.com/technology/2023/may/02/geoffrey-hinton-godfather-of-ai-quits-google-warns-dangers-of-machine-learning>.
  - [78] C. Kang, Openai's Sam Altman urges a.i. regulation in senate hearing, 2023. URL: <https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html>.
  - [79] A. Belanger, "meaningful harm" from ai necessary before regulation, says microsoft exec, 2023. URL: <https://arstechnica.com/tech-policy/2023/05/meaningful-harm-from-ai-necessary-before-regulation-says-microsoft-exec/>.
  - [80] C. Hetzner, Ex google ceo wants a.i. regulation left up to big tech, 2023. URL: <https://finance.yahoo.com/news/former-google-ceo-eric-schmidt-155901279.html>.
  - [81] L. Lopez, Ai is silicon valley's desperate, last-ditch attempt to avoid a stock market wipeout, 2023. URL: <https://www.businessinsider.com/ai-technology-chatgpt-silicon-valley-save-business-stock-market-jobs-2023-5>.
  - [82] A. Bharade, A widow is accusing an ai chatbot of being a reason her husband killed himself, 2023. URL: <https://www.businessinsider.com/widow-accuses-ai-chatbot-reason-husband-kill-himself-2023-4>.
  - [83] G. A. Fowler, Snapchat tried to make a safe ai. it chats with me about booze and sex, 2023. URL: <https://www.washingtonpost.com/technology/2023/03/14/snapchat-myai>.
  - [84] B. Perrigo, Bing's ai is threatening users. that's no laughing matter, 2023. URL: <https://time.com/6256529/bing-openai-chatgpt-danger-alignment/>.
  - [85] R. A. Crof, Benkler, y. (2006). the wealth of networks: How social production transforms markets and freedom. new haven and london: Yale university press. 528 pp., Social Science Computer Review 26 (2008) 259–261. doi:10.1177/1084713807301373.
  - [86] C. DiBona, S. Ockman, Open sources: Voices from the open source revolution, " O'Reilly Media, Inc.", 1999.

- [87] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, in: *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., 2019, pp. 8024–8035. URL: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [88] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, et al., Tensorflow: A system for large-scale machine learning, in: *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.
- [89] I. Solaiman, The gradient of generative ai release: Methods and considerations, in: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, FAccT '23*, Association for Computing Machinery, New York, NY, USA, 2023, p. 111–122. doi:10.1145/3593013.3593981.
- [90] P. Hacker, A. Engel, M. Mauer, Regulating chatgpt and other large generative ai models, in: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, FAccT '23*, Association for Computing Machinery, New York, NY, USA, 2023, p. 1112–1123. doi:10.1145/3593013.3594067.
- [91] Y. Yakimova, Ai Act: a step closer to the first rules on artificial intelligence | news | european parliament, 2023. URL: <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>.
- [92] L. Bertuzzi, Ai act moves ahead in eu parliament with key committee vote, 2023. URL: <https://www.euractiv.com/section/artificial-intelligence/news/ai-act-moves-ahead-in-eu-parliament-with-key-committee-vote/>.
- [93] H. Face, The bigscience rail license, 2022. URL: <https://bigscience.huggingface.co/blog/the-bigscience-rail-license>.
- [94] J. Cobbe, M. Veale, J. Singh, Understanding accountability in algorithmic supply chains, in: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, FAccT '23*, Association for Computing Machinery, New York, NY, USA, 2023, p. 1186–1197. doi:10.1145/3593013.3594073.
- [95] E. Parliament, Council, Draft compromise amendments on the draft report: Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts - general approach, 2023-05-16. URL: [https://www.europarl.europa.eu/meetdocs/2014\\_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA\\_IMCOLIBE\\_AI\\_ACT\\_EN.pdf](https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf).
- [96] J. Baudrillard, *Simulacra and simulation*, University of Michigan press, 1994.
- [97] E. Hutchins, Cognitive ecology, *Topics in cognitive science* 2 (2010) 705–715. doi:10.1111/j.1756-8765.2010.01089.x.
- [98] G. Lakoff, *Don't think of an elephant!: Know your values and frame the debate*, Chelsea Green Publishing, 2014.
- [99] E. Parliament, Council, Proposal for a directive of the european parliament and of the council on adapting non-contractual civil liability rules to artificial intelligence (ai liability directive), 2022-09-28. URL: [https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence\\_en](https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en).

- [100] T. H. Davenport, J. C. Beck, The attention economy, *Ubiquity* 2001 (2001) 6. doi:10.1145/376625.376626.
- [101] F. Ferrari, S. Moruzzi, *Verità e post-verità: dall'indagine alla post-indagine*, 1088press, 2020.