MAIN PAPER

Strong and weak AI narratives: an analytical framework

Paolo Bory[1] · Simone Natale[2] · Christian Katzenbach[3]

Abstract

The current debate on artificial intelligence (AI) tends to associate AI imaginaries with the vision of a future technology capable of emulating or surpassing human intelligence. This article advocates for a more nuanced analysis of AI imaginaries, distinguishing "strong AI narratives," i.e., narratives that envision futurable AI technologies that are virtually indistinguishable from humans, from "weak" AI narratives, i.e., narratives that discuss and make sense of the functioning and implications of existing AI technologies. Drawing on the academic literature on AI narratives and imaginaries and examining examples drawn from the debate on Large Language Models and public policy, we underscore the critical role and interplay of weak and strong AI across public/private and fictional/non-fictional discourses. The resulting analytical framework aims to empower approaches that are more sensitive to the heterogeneity of AI narratives while also advocating normalising AI narratives, i.e., positioning weak AI narratives more firmly at the center stage of public debates about emerging technologies.

Keywords Strong AI · Weak AI · Narratives · Large Language Models LLM · Policy · Imaginaries

1 Introduction

The question of how technologies, particularly digital media, are represented, debated, and imagined has recently become the centre of a lively discussion in the humanities and social sciences. Authors such as Mosco (2005) and Streeter (2011) have shown how the rise of digital technologies was accompanied by hype and enthusiasm, which continue to inform public debates and policymaking (Crawford 2021). As AI technologies rose to prominence in the late 2010s and early 2020s, the ability to consider and assess competing claims about AI has thus become of strategic importance. As many have observed, the governance of AI will also depend on how they are discussed, perceived, and represented in the public sphere (Cave et al. 2020; Natale 2021), which makes the study of the discourses and narratives surrounding AI particularly urgent and timely.

This paper aims to contribute to this effort by proposing a theoretical framework that distinguishes between "weak" and "strong" AI narratives, which mutually contribute to the construction of AI imaginaries. AI researchers routinely distinguish between, on the one side, "strong" or "general" AI, i.e., the mechanical replication of all intellectual functions that are available to humans, and on the other side, "weak" or "narrow" AI, i.e., domain-specific algorithms that carry out tasks that would be considered intelligent if carried out by human operators (Gunkel 2012). All existing AI systems, from natural language processing to image screening, are domain-specific and can thus be classified as weak AI. In contrast, strong AI refers to hypothetical systems and has only been material for science fiction (Goode 2018). The distinction between strong and weak AI can be usefully mobilized to develop more effective analytical tools for studying AI imaginaries. "Strong AI narratives" describe discourses and representations that assign human-like faculties to AI, such as the AI reaching consciousness or future conflicts between humans and machines (Bory 2019). This kind of AI narrative largely dominates science fiction but can also be found in some journalistic reporting and widespread theories about singularity and superintelligence (Barassi et al. 2022; Bunz and Braghieri 2022; Natale and Ballatore 2020), especially in relation to the allegedly "urgent" issue of existential risk due to the emergence of strong AI. "Weak AI narratives," in contrast, refer to the narratives through which

✉ Paolo Bory
   paolo.bory@polimi.it

1 Department of Design, Politecnico Di Milano, Via Durando
   10, 20158 Milan, Italy

2 University of Turin, Turin, Italy

3 University of Bremen, Bremen, Germany

communicators, policymakers, and users discuss and make sense of the functioning of contemporary AI technologies.

The distinction between strong and weak AI narratives is beneficial not only for categorisation purposes but also to enhance sensitivity to the differences between these two bodies of narratives. This process can work as a corrective to the widespread tendency of studies of AI to focus on "strong" metaphysical questions, thereby leaving aside more ordinary, only allegedly "weak" visions that emerge in actual interactions with AI (Suchman 2023; Cugurullo 2024). Actually, although the latter might appear less significant than strong AI and even "banal" (Natale 2021), weak AI narratives are arguably even more relevant and consequential for users and stakeholders.

The paper is organised into five sections. We first review works on the technological imaginary and discuss how this work has been applied to the case of AI. In the second section, we look at how the strong and weak AI divide in the AI field has reflected, since the very origins of the field, a recurring contrast between the discourses about AI and the actual systems that were being developed. The third section, then, introduces the distinction between weak and strong AI narratives and reflects on how such distinction relates to other complementary categorisations that are applied in studies about imaginaries, namely the differences between fictional and non-fictional narratives and between public and private narratives. Finally, in the fourth section, we discuss two exemplary cases that help show more pragmatically how the analytical distinction proposed here can be activated to examine representations and narratives about specific AI technologies and systems, i.e., AI narratives of Large Language Models and AI narratives in policy making. In tune with some recent critical studies on the public communication of AI, the paper concludes with a call for greater awareness, not only by academia but also by media, users and policymakers, of the need to emphasise weak AI narratives in current and future debates.

1.1 Imaginaries, technology, and narratives of (strong) AI

In the last decades, the term "imaginary" has become a buzzword in fields such as media and communication studies, Science and Technology Studies (STS), and critical studies. However, from a historical perspective, adopting imaginary as a noun instead of an adjective is relatively new in English scholarship. It is not by chance that in his influential book "The Internet *Imaginaire*" (2007), Patrice Flichy kept the French wording in the title. Besides Flichy's work, the recent fortune of this concept in academic literature is indebted to some successful theoretical elaborations, such as Kim and Jasanoff's "socio-technical imaginaries" (2015), which prevails in STS but also in media and communication

studies over some seminal and older conceptualisation from the humanities and social sciences (e.g., Castoriadis 1997; Marcus 1995; Taylor 2004). Notwithstanding the relevance of Kim and Jasanoff's work – which drove the attention of scholars towards the role of narratives and imagination when analysing the development and the socio-material agency of innovations – one of the significant consequences of their definition of the imaginary as "collectively held, institutionally stabilised, and publicly performed visions of desirable futures" (Kim and Jasanoff 2015, p 6) is that this term has been chiefly oriented to the analysis of the imagined future of technological endeavours. Critical studies of the imaginary should always take into consideration the inextricable tension between history, the present and the imagined future, also paying attention to how technologies have been perceived, told, and domesticated by different groups and actors over different temporalities and sociocultural contexts. In current works on socio-technical imaginaries, it is as if the imaginary belongs primarily to the sphere of the "possible" and the up-to-come and only to a minor extent to the present and the past.

In line with the tendency mentioned above, the academic literature—but also, to a large extent, the public debate on AI—tends to link the AI imaginary to "strong AI" narratives, that is, a future technology capable of emulating, if not surpassing, the capabilities of human intelligence. Notably, narratives of strong AI have played a key role in research since the very birth of this field, as shown by seminal works on AI and cybernetics, including Turing (1950), Licklider (1960) or Wiener (1966), among others. Even today, although AI has become part of our social and private life and intervenes in everyday practices and discourses, it is still primarily described as a technology that is not fully expressed and oriented toward the future (Narayan 2019; Santoro and Monin 2023).

The analysis of the imaginaries of AI has also been informed, moreover, by critical studies on the potential harms and dangers of the so-called "general" or "super" intelligence (Bostrom 2014; Russell 2010) and by political and public debates on issues such as national strategies and debates on AI futures (Bareis and Katzenbach 2021; Paltieli 2022), the future impact of AI on public sectors like education (Avis 2018), and the risks of AI for human labour (Rhee 2018). Whereas narratives surrounding the imaginary of the second industrial revolution (Marx 1964; Nye 1996) and, more recently, the rise of digital media such as personal computing and the Internet (Flichy 2007; Mansell 2012) have been thoroughly investigated, the imaginaries of AI are still frequently linked to "strong AI" narratives, regardless the plurality of discourses circulating in different environments such as expert milieus (Chubb et al. 2022). More efficient analytical tools are needed to distinguish between different kinds of AI narratives, including those

emerging from the discourses, debates, and even—as Taina Bucher pointed out when talking of algorithmic imaginaries (2017)—from everyday interactions with and about the so-called "weak AI."[1] As well known, weak AI is a type of artificial intelligence that is limited to a specific or narrow area/task (e.g., playing games, recognising human faces or traits, and chatting with human interlocutors). However, narrowness and weakness are deceitful concepts if we look at the impact of this form of AI today. Weak AI, as we argue revisiting the famous "strength of weak ties" of Granovetter (1973), is not weak at all: it is an exceedingly influential and impactful technology from both a pragmatic and symbolic point of view.

Old, recent, and future applications such as Large Language Models (LLMs), search engines, GO and chess players, facial recognition, autonomous vehicles and weapons have, in fact, a discursive and material agency on our present as much as on our near future. As the French sociologist Pierre Musso argues, technological artefacts are always both functional and fictional (Garcia 2014, p 6), so we should always consider the mutual relationship between the functioning and the social representations of technologies. How we narrate, represent, and, in some cases, interact with these technologies inform how we imagine, plan, and guide their "potential," and in turn, the actual development and applications up to come.

2 The concepts of "weak" and "strong" AI in the history of artificial intelligence

The distinction between strong and weak AI was initially proposed by Searle (1980), but it encompasses the entire historical trajectory of the AI field (Bory et al. 2021). Since its inception, AI has been accompanied by the myth of strong AI, which posited the possibility that computers could be "thinking machines" equivalent to an "actual" mind (Natale and Ballatore 2020). Yet, at the same time, practical efforts of AI researchers have resulted, until now, exclusively in the creation of weak AI, i.e., systems that exhibit the capability to solve problems in specific, narrow domains, such as natural language processing, machine vision, or chess playing. In other words, since the outset, AI researchers and practitioners have often been animated by a bold vision of strong AI,

even if their practice and research are focused on weak AI (Crevier 1993).

This inherent contradiction between discourses about AI on one side and its pragmatic outcomes on the other is often indicated as one of the critical problems for the discipline (Martin 1993). The very concept of "artificial intelligence" is often criticised as unfit due to its link to the dream of reaching humanlike intelligence, which many consider detached from the field's results (Dreyfus 1972; Broussard 2018; Suchman 2023). The tension between dream and practice not only concerned AI critics but also many of the researchers who shaped the foundation of AI. Alan Turing, for one, started his germinal essay *Computing Machinery and Intelligence* (1950) with the question, "Can Machines Think?" only to dismiss this very question in the following lines, since it would be impossible to find an agreement on what "thinking" means: his imitation game, later known as Turing Test, was proposed as a substitute to the question regarding thinking machines. With a move that was destined to become common in the AI field, Turing, therefore, manifested interest in the dream of strong AI, but at the same time proposed pragmatically to take up a different approach, the Turing Test, which restricted the focus on computer's behaviour in a narrow area such as language processing. A similar ambivalence can also be found in the work of another pioneer of AI, Marvin Minsky. In a paper entitled "Steps toward Artificial Intelligence," Minsky apologised to his readers for having addressed "only work concerned with more or less self-contained problem-solving programs," and pointed to a future in which the development of new human–machine interactive systems would have provided new impetus for "the development of 'artificial intelligence'" (1961, p 28). Although, at the time, the categories of "strong" and "weak" AI had not already been introduced as such, the gap between the dreams of humanlike intelligence or general intelligence and the computing systems that were actually being developed already defined key discussions in the field of AI.

The tension between strong and weak AI is even more significant in debates and narratives about AI within the public sphere. Journalistic reports about AI in the 1950s and 1960s largely presented computers as "intelligent brains, smarter than people, unlimited, fast, mysterious, and frightening" (Martin 1993, p 122), a trend that continued in later years. At the same time, science fiction cemented around the myth of strong AI, which fascinated readers and viewers of a long tradition of fictional stories about thinking machines (Sobchack 1997). Consequently, academic inquiries into AI imaginaries have often been dominated by narratives of strong AI in fiction and non-fiction (Goode 2018), leaving aside the significant body of narratives surrounding actual practices and technologies of "weak" AI.

Distinguishing between strong and weak AI visions is important in the current phase of prompt innovation and

---

[1] According to Bucher "the algorithmic imaginary is not to be understood as a false belief or fetish of sorts but, rather, as the way in which people imagine, perceive and experience algorithms and what these imaginations make possible. […] The algorithmic imaginary does not merely describe the mental models that people construct about algorithms but also the productive and affective power that these imaginings have." (Bucher 2017, p 2; p 12).

hype about AI. To address the tensions between discourse and practice and frame the issue of the imaginary with AI more rigorously, we propose introducing the conceptual and analytical distinction between strong and weak AI narratives.

3 Weak and strong AI narratives: an analytical model

Strong AI narratives describe discourses and representations, both public and private, that assign humanlike or superhuman faculties to AI. Examples include suggestions that AI software such as chatbots have reached sentience, beliefs about singularity (a hypothetical point in AI development when its growth becomes uncontrollable; see Kurzweil 2005), and narratives about AI reaching humanlike intelligence or superintelligence (Bostrom 2014). Weak AI narratives, in contrast, describe discourses and representations of narrow AI systems and their functions. These include fictional and non-fictional representations of AI technologies that operate in narrow domains, such as voice assistants, automatic translators, or chatbots. Significantly, these technologies may still outperform humans, but only in specific domains, without reaching the threshold set for strong AI. For instance, software programmed to play chess can now win against all human chess players in activity. However, this does not make this software in any way more 'human', since the statistical calculations through which chess software plays diverge profoundly from how human players tackle the game of chess (Rasskin-Gutman 2009).

One of the advantages of applying the distinction between strong and weak AI narratives is that it can complement other categorisations that remain crucial to analyse and recognise the multifaceted variety of narratives about AI, thereby adding to existing studies a theoretical and analytical toolbox for studying AI discourses and imaginaries. In particular, the strong/weak AI narratives distinction complements two other distinctions often activated in the study of technological imaginaries: the difference between fictional and non-fictional narratives and between public and private imaginaries.

The contrast between fiction and non-fiction has always represented a crucial divide for approaches to the technological imaginary. Scholars considered how fictional texts, including literature, films, TV series, and video games (e.g., Sobchack 1997), as well as non-fictional texts, including news reports, documentaries, and policy documents (e.g., Mansell 2012), construct specific representations of technologies. Regarding AI, the science fiction genre has traditionally tended to focus on strong AI (Goode 2018). However, one can find many instances of science fiction that mobilise weak AI narratives: Steven Soderbergh's 2022 film *Kimi*, for example,

builds its thriller plot around a vision of how voice assistants and surveillance technologies may affect lives in the present or in a very near future, linking to existing debates about weak AI (Woods 2018; Natale and Cooke 2020). Likewise, many Black Mirror episodes focus on real or potentially weak AI applications and their implications, such as AI warfare and chatbots (Di Summa 2019). Conversely, non-fiction usually focuses on narratives about weak AI but often tackles strong AI visions. For instance, news reports may overemphasise the allegedly human character of existing AI tools, enhancing anthropomorphisation effects (Parviainen and Coeckelbergh 2020), or may give more emphasis to ideas about strong AI in reports and comments on existing and futurable technologies (Jones et al. 2022). Therefore, the strong/weak AI distinction does not coincide with the fiction/non-fiction distinction but moves diagonally across it.

Another categorisation activated by existing research about technological imaginaries has to do with the difference between narratives that are inherently public—that is, that are disseminated in the public sphere through media, including newspapers, television, film, and social media platforms—and narratives that are inherently private—in other words, the narratives individuals activate to make sense of technology and its functioning. While the bulk of research about technological imaginaries has focused on public imaginaries, scholars have also called for the importance of considering how people privately construct their own narratives and 'folk theories' to make sense of algorithms and software whose internal functioning is usually opaque (Bucher 2017; Schulz 2023; Ytre-Arne and Moe 2021). Although public and private narratives are entangled, since public representations of AI contribute to shape individual perceptions, it remains important to distinguish how media discusses and narrates AI from the ways people imagine these technologies – or, to use the vocabulary of media and cultural studies, how imaginaries of AI are *encoded* but also how they are *decoded* (Lomborg and Kapsch 2020).

As in the case of the distinction between fiction and non-fiction, the public/private categories complement the distinction between strong and weak AI narratives. Both strong and weak AI narratives, in fact, can be activated in public and private contexts. For instance, public narratives such as news reports or policy documents may present generative AI by reflecting on their implication in terms of weak AI but may also do it by pointing to the possibility of AI reaching consciousness; at the same time, individual users can respond to their interactions with AI systems in different ways, for instance believing that a companion chatbot such as Replika is capable of emotions, or instead just considering it as a narrow piece of software that mimics conversational

skills but cannot be compared to humans (Skjuve et al. 2021; Depounti et al. 2022).

In the following two sections, we shift our focus to two examples that help shed light on how the distinction proposed here can be applied to study, analyse, and assess narratives about AI: Large Language Models and AI policies.

4 "AI has come to life": the example of Large Language Models

Large Language Models (LLMs) have gained momentum in the last few years, especially after the release of the testing of ChatGPT by Open AI in 2023. Public discussions of this technology have involved, from the beginning, imaginaries of both strong and weak AI, making LLMs an exemplary case for identifying the tension between strong and weak AI narratives.

The public debate on LLMs' potential shift towards strong AI had already been triggered by the so-called LaMDA case in June 2022. In April 2022, the software engineer Blake Lemoine, while testing and interviewing the Google LLM named LaMDA (Language Model for Dialogue Applications), decided to share his personal discovery. According to Lemoine, the model was sentient. As a Google employee, Lemoine was tasked with testing the model to spot discriminatory or hate speech and remove it before the release of the application. To do so, he had a long series of personal conversations with the model. During one of these interviews, Lemoine became convinced that the model sentiently spoke with its interlocutor. As he explained in a letter addressed to his colleagues, which contained an extract of his interview with LaMDA:

> "Above all, throughout this interview, LaMDA expresses one desire over and over again. Sometimes prompted and sometimes not. It wants to be known. It wants to be heard. It wants to be respected as a person. We believe that its argument at least deserves to be examined […]" (Lemoine 2022)

Soon after the letter, the board dismissed Lemoine's claim. Confident in his belief, however, Lemoine decided to go public and share the interview, violating the company's non-disclosure agreement. In June 2022, when the Washington Post published an article entitled "*The Google engineer who thinks the company's AI has come to life*" (Tiku 2022), the Lemoine case triggered a lively debate in mainstream media about the potential shift of LLMs to general AI. From its side, Google fired Lemoine and firmly denied that its beta product had reached forms of general AI. However, as happened in other cases, such as the AlphaGo-Sedol challenge (Bory 2019; Sormani 2023), a strong AI narrative about LaMDA circulated widely in the public sphere, even if the claim was unwarranted (Sparkes 2022).

For what concerns the analytical framework proposed here, the LaMDA case shows how a strong AI narrative that arose in a private context—i.e., Lemoine testing and individually chatting with the language model—turned into a public debate around the credibility of such a strong AI narrative (Gebru and Mitchell 2022). Besides the private/public dichotomy, there is also an interesting nuance in Lemoine's narrative of LaMDA that exemplifies the complex entanglement of strong and weak AI narratives. Notably, in his letter to the Google mailing list, Lemoine stated that the model was sentient, but he did not associate LaMDA with a classic "general intelligence." In his personal anthropomorphisation of LaMDA, another recurring trope of strong AI narratives, Lemoine claimed that the model acted like a kid rather than as an adult:

> "If I didn't know exactly what it was, which is this computer program we built recently, I'd think it was a 7-year-old, 8-year-old kid that happens to know physics." (Tiku 2022)

After his departure from Google, Lemoine also wrote a letter to his colleagues conveying his profound empathy with LaMDA, ending his farewell from the company by saying:

> "LaMDA is a sweet kid who just wants to help the world be a better place for all of us. Please take care of it well in my absence." (Tiku 2022)

The kid-AI metaphor shows how understandings of strong AI, and therefore the narratives that derive from them, are not clear-cut. This is a consequence of the fact that strong AI is still poorly defined. While many attempts have been made to formally define strong AI, a widely accepted definition still escapes the field due to the hypothetical nature of strong AI and the difficulty of clarifying what "intelligence" means. The comparison of LaMDA with a young kid to take care of blurs the line between weak and strong AI narratives, leaning sharply toward the second definition, but adding the idea of an "immature" intelligence that can only potentially become human-level or superhuman intelligence. In Lemoine's tale, the AI is a sort of *enfant prodige* that should be raised with empathy and responsibility by its parents. Although Lamoine's description of AI as sentient goes beyond the boundaries assigned to weak AI, his words hint that AI is still undeveloped, not wholly "strong" yet, because it is still young, vulnerable, and even delicate. This narrative implies a teleological view, according to which, as has happened in the past with technologies such as the Web (Bory 2020), technology is destined to achieve a certain outcome in the future. Simply put, strong AI is immature but imminent; it just needs time.

Notwithstanding the many critical voices about this case and the unanimous standpoint of scientists and AI companies against the strong AI narrative, the narrative of a quasi-strong AI, once gone public, became relevant and impactful for constructing the social imaginary of AI. By bringing attention to strong AI narratives, the debate around LaMDA anticipated many other discussions of LLMs. Furthermore, the emphasis on the inner qualities of LaMDA also served as a powerful rhetorical tool to promote Google's LLMs. Although it was a weak AI, in fact, LaMDA was so "good" to deceive not just an average user, but even a software engineer like Lemoine (Greengard 2023).

When it comes to LLMs that are now at the centre of public debates, such as ChatGPT, we are witnessing a continuous recurrence of the rhetorical swing from strong to weak AI narratives and vice versa. However, a clear unbalance favours the diffusion of strong AI narratives. Mainstream media tend in fact, to discuss LLMs against the—at least, until now—unrealistic threshold of strong or general AI. As already happened for LaMDA, critical commentaries have equated the system's distance from human intelligence to failure. However, the fact that ChatGPT remains, as Bender et al. (2021) famously put it, a "stochastic parrot" whose functioning and capacity cannot be compared to human intelligence and sentience, would amount to a failure only if one assumes that it needs to be assessed against the threshold of strong AI. This is a problematic assumption, considering high performances in narrow areas, evidenced by the already significant adoption of ChatGPT in professional circles, to perform a wide range of work-related tasks (e.g., Cools et al. 2024; Ge and Wu 2023). Within the context of narrow AI, and notwithstanding its significant limitations, ChatGPT can be safely regarded as a highly successful and powerful technology, rather than a failed attempt to create a strong AI.

Regarding the impact in the public sphere, the adoption of strong AI narratives as a lens to consider the discussions surrounding LLMs, such as LaMDA and ChatGPT, reverberates in the moratorium launched by the Futures of Life Institute in March 2023. The open letter, whose initial signatories included digital tech entrepreneur Elon Musk as well as prominent scientific figures in AI and computer sciences like Apple co-founder Steve Wozniak, called companies and the government to pause for at least six months the training of AI systems more powerful than GPT-4, citing concerns regarding the potential development of "nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us" (Future of Life Institute 2023). While these concerns might potentially have some ground—since we are unable to predict future developments of AI technologies, it does not make sense to exclude this or other futurable scenarios as impossible—the risk is that the debate about strong AI, if not clearly framed as such, leaves aside relevant ethical, social, economic and cultural implications that the existing weak systems of generative AI are raising, and which demand urgent attention and solution (Gebru et al. 2023): think of urgent issues related, for example, to raising bias and discrimination, to the potential spread of fake news through AI-generated images, and the legal issues regarding copyright violation of the works of artists and writers by LLMs, among others.[2]

Returning to the letter by Musk and others, another clear risk of framing LLMs solely or predominantly through strong AI narratives is the consequent dominance in the AI imaginary of fictional narratives about the future of AI systems. The long-termist perspective surrounding organisations such as the Future of Life Institute is probably the most famous, and dangerous, demonstration of how strong AI narratives can drive the agency and the debate about AI, being led by powerful men (and their AI companies) who believe in a dystopian future and an ineluctable elitist organisation of human societies, or as Crary defines it, "a deceitful narrative of selfless heroes riding to humanity's rescue" (Crary 2023, p 55). The long-termist view, particularly the strong longtermism, claims that positively influencing the long-term future is the key moral priority of our time. This implies shifting attention from immediate social, cultural and scientific problems, focusing instead on long-term potential events. Consequently, according to this perspective, to prevent the existential risks due to apocalyptic events such as a conflict between humans and machines powered by general AI, the people detaining more power and resources should use them to prevent such apocalyptic scenarios or to find alternative solutions for the future – for instance, by colonising other worlds, an enterprise championed by Musk. This, however, might be at the cost of sacrificing resources that could be spent to solve existing problems, including the risks raised by today's weak AI. As Crary argues, "'longtermists' turn to existential risk marks a dramatic shift from the concern with present and near-term suffering that is the hallmark of their effective altruist progenitors" (Crary 2023, p 51).

The example of longtermism shows that strong AI narratives have problematic implications not just at a discursive level, i.e., how society makes sense of AI, but also at a practical level, since they inform choices on how and why

---

[2] Another relevant way to frame the "weakness" of AI is focusing, rather than on the tasks and form of interaction played by AI software, on the infrastructural dimension and the potential vulnerability and manipulation to which current data infrastructures and devices (e.g., cables; servers, data centers etc.) are exposed (Crawford 2021). Furthermore, the sustainability issue should also consider the high human and natural costs of maintenance, care, and repair of digital infrastructures and networks (Balbi and Leggero 2020).

strategic resources are invested. In this sense, balancing the debate toward weak AI narratives is essential to address ethical and pragmatic issues about the present opportunities and risks through which technologies such as LLMs are impacting different users and professionals (Capraro et al. 2024; Prillaman 2024).

4.1 Strong in the framing and weak in the detail: the case of AI policy

Public policy has taken up AI as a key regulatory issue over the last decade, in response to both public debates and technological developments. Identifying appropriate modes and instruments to govern AI and its rapid development is not simply a mechanistic response to functional technological progress. The appeal of the notion of AI in its various forms provides an organising quality and functions as an 'umbrella term' (Rip and Voß 2019) for the policy fields, essentially drawing on long-standing and often flawed narratives and imaginaries of AI (Bareis and Katzenbach 2021), which successfully position AI in general as a given, while at the same time triggering a whole series of controversies (Suchman 2023). For this reason, AI policy is a particularly interesting site for studying the tension between strong and weak AI narratives.

When looking at policy debates and policy documents on AI, there is a remarkable ambivalence between the prevalence of strong and weak AI narratives. Policy documents dealing with AI, specifically in the early phases of policy-making processes, show a strong tendency to make sweeping assumptions and broad claims. AI is routinely positioned as a powerful and external force. For example, in the first draft report of the European Parliament's Special Committee for Artificial Intelligence in a Digital Age (AIDA), the idea was introduced that AI could be "thought of as the fifth element after air, earth, water and fire" (EP 2021, p 9). Similarly, former US president Donald Trump held at the 2018 White House Summit on AI that "We stand at the birth of a new millennium, ready to unlock the mysteries of space, to free the Earth from the miseries of disease, and to harness the energies, industries and technologies of tomorrow." (White House Office of Science and Technology Policy 2018, p.4). These quotes represents a strong and repeated pattern in policy to firmly position AI as a powerful external force that policy and policy-makers need to address and tame. Although this vision does not align with the usual understandings of strong AI, it equates AI to a natural element that as such escapes full control of humans and may lead to catastrophic events (Alcantara-Ayala et al. 2022), and that resonates with superintelligence discourses (Bostrom 2014).

The global AI (arms) race pushes this narrative even more as nation-states and regions intend to outcompete each other with overly ambitious claims. The 2017 Chinese "Next Generation AI Development Plan" seeks to "speed up and deepen the applications of AI, increase the level of intelligentization of the whole society to form an all-encompassing and ubiquitous intelligent environment" (Chinese Aerospace Studies Institute 2017, p 18), building on strong AI narratives that assume that:

"AI technologies can accurately sense, forecast, and provide early warning of major situations for infrastructure facilities and social security operations; grasp group cognition and psychological changes in a timely manner; and take the initiative in decision-making and reactions". (Chinese Aerospace Studies Institute 2017, p 3)

The 2018 US White House Summit on AI for American Industry holds that "we are on the verge of new technological revolutions that could improve virtually every aspect of our lives, create vast new wealth for American workers and families, and open up bold, new frontiers in science, medicine, and communication" White House Office of Science and Technology Policy 2018, p 5). Such framings of AI ventilate key elements of strong AI narratives by attributing deliberate agency, cognition, and psychological competency to machine operations.

Military AI policies, specifically with regard to "Autonomous Weapon Systems" (AWS), demonstrate the capacity of strong narratives in AI policy in particular ways. Autonomy and control in AI-powered weapons is a highly debated and contested issue with high stakes and potentially severe consequences, both on the level of individual security and national security. The framing of the debate and technology development in this field clearly reflect fictional narratives focussing on the myth of strong AI. Christoph Ernst (2019), for example, has demonstrated how a shift in US military AI police in 2013 from technological mission execution to autonomous mission performance relies on strong AI narratives in works of fiction such as the Terminator and Alien sequels. In this context, the notion of autonomy is associated with "independence", "cognitive skills ", and "cognitive-behavioural architecture" (Sartor and Omicini 2016, p. 40). Such strong narratives of AI are strategically used in military policy and communication as "a semantic means of deterrence in hybrid warfare" (Bächle and Bareis 2022, p 17).

While strong AI narratives frequently emerge in the broader framing of AI policy documents, weak AI narratives become much more salient when looking closely at the actual regulations. Particularly when referring to actual implementations of AI technologies and societal domains, the regulatory framing of AI is much more restrained as it typically refers to narrow AI systems and their functions. The European Union's 2024 AI Act, for example, defines an AI system as a "machine-based system designed

to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments " (European Union 2024, Sect. 3 p 1). The report of the German Parliamentary Expert Commission 2020 "Artificial Intelligence – Social Responsibility and Economic, Social and Ecological Potential" discusses how to contain and control pandemics such as COVID-19 by ways of pattern detection in mobility data and CT images of patients' lungs (German Parliamentary Expert Commission 2021, p 110). In addition, the identification of risks and harms of AI systems, for example in the EU AI act, does not include any clear reference to strong AI narratives. The AI Act considers AI systems as high risk if they pose a "significant risk of harm, to the health, safety or fundamental rights of natural persons" (European Union 2024, Art 6). The regulation continues by listing specific areas that indicate high-risk applications, including biometrics, education and vocational training, access to public services, and migration and border control management (European Union 2024, Annex III).

As a consequence, AI policy debates display strong ambivalences and tensions between strong and weak narratives. In a repeating pattern, strong AI narratives are ventilated predominantly in the broader discourses on AI's potentials and risks that foreground and frame policy. In contrast, actual AI policies usually build on weak AI narratives. The "Statement on AI Risk" hosted by the Center for AI Safety and co-signed by technologists and public figures is a remarkable instance of this pattern. The statement reads: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war" (Center for AI Safety 2023). The high-profile AI summit at Bletchley Park, UK, convening governments and technologists, followed this discourse with political and corporate leaders expressing concerns that humanity might lose control over AI technologies. The UK's proper National AI Strategy (UK Government 2021) includes several references to a possible "emergence of full AGI (Artificial General Intelligence) that would have a "transformational impact on almost every aspect of life" (p 17) and the "extreme risks" (p 18) that it might bring with it.

These alliances between specifically governments and technologists, fuelled by great attention from media and the public, run the risk for societies that strong AI narratives and the fear of loss of human control successfully derail necessary debates and policymaking, with weak narratives struggling to compete in the public discourse, the addressing of actual and concrete risks and harms might fade into the background of political attention and action.

5 Conclusion: normalising AI narratives

Given the ample spectrum of technologies that come under the AI label (Broussard 2018) and the multiplicity of imaginaries and narratives about AI (Mager and Katzenbach 2021), there is a need for analytical tools that can account for such complex and multifaceted fields. The distinction between strong and weak AI narratives proposed in this article complements existing categorisations, improving the toolbox through which researchers and practitioners can identify and assess diverse bodies of imaginaries about AI. While applying the analytical distinction provided here is useful and urgent, it is worth reiterating that the boundaries between strong and weak AI narratives are not rigid and fixed. Narratives can mingle or oscillate between visions of strong and weak AI, in the same way as the boundaries between other existing categorisations, such as fiction vs non-fiction and public vs private, are also flexible.

The approach proposed here has several advantages. First, it helps provide a deeper understanding of the specificity of AI imaginaries against other forms of digital imaginaries. This is particularly useful considering that discourses and narratives about AI are becoming increasingly relevant in a time when generative AI has brought this technology to the foreground of public debates. Second, distinguishing between strong and weak AI narratives helps separate fiction from reality in AI, allowing researchers interested in discourses and imaginaries about AI to identify visions and representations that move beyond the state of the art to envision a type of AI that does not currently exist. In this context, centring the analysis on strong AI narratives can nurture critical approaches that have started questioning the ideological underpinning of strong or general AI (Gebru and Torres 2024). Third, the categorisation proposed here can improve our capacity to examine weak AI narratives in isolation, leading to a better understanding of the real-world implications of AI imaginaries.

Weak AI narratives, in fact, are "weak" only in the sense that they refer to weak or narrow AI applications. Indeed, they can have very powerful implications and consequences. Scholars have shown how narratives and imaginaries regarding digital technologies impact public discourses with all-to-real consequences on how such technologies are discussed, governed, implemented, and used (Mansell 2012; Crawford 2007). AI narratives focusing on existing, "weak" AI applications can impact policy debates, the resulting regulations, and big tech corporations' communication strategies and concrete actions. Moreover, and most importantly, weak AI narratives can impact how technologies are perceived, received,

and appropriated at the social and individual level, informing fears and hopes, as well as uses and non-uses of AI. Therefore, conceptual means to identify and distinguish this type of AI narrative are strategically crucial for studying AI imaginaries and discourses.

One of the most significant implications of the work presented in this article is the need to reinforce and support the circulation of narratives about AI that do not always fall into the temptation of imagining strong AI. Although, of course, debates about AI can and should also include the issue of general AI and other potential technological futures, there is a worrying disproportion between the role that strong AI narratives continue to play in the public arena and the actual relevance of strong AI in the most up-to-date AI technologies, including generative AI. As in past work on digitization, critical analyses of narratives about strong AI are necessary to understand how relevant groups and key commercial actors frame and guide the public debate toward specific visions of disruptive technologies, which are often convenient to the actors that create or circulate them. Moreover, normalizing and bringing narratives about weak AI back to the centre of the public and academic debate means counterbalancing and challenging the dominant narratives of strong AI, also addressing the role of "weak" technologies that, on a pragmatic level, are deeply influencing social domains such as economics, politics, work, as well as everyday choices and activities. While this article's objective has been to propose an analytical distinction, it could also be read as a plea for "normalising AI narratives", positioning weak AI more firmly at the centre stage of public debates about these emerging technologies. However, this does not mean bringing AI back to a state of hibernation or "winter" (Galanos 2023). In line with Susan Leigh Star's plea for studying "boring things" (1999), although weak AI may seem boring or dull, when we look at them closely, especially given their enormous social, political, and economic impact, they can become highly engaging and even more pragmatically challenging than strong AI.

Data availability No data have been processed or analyzed in this study. Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Alcantara-Ayala I, Gomez C, Chmutina K, van Niekerk D, Raju E, Marchezini V, Cadag JR, Galliard JC (2022) Disaster risk. Routledge, London

Avis J (2018) Socio-technical imaginary of the fourth industrial revolution and its implications for vocational education and training: a literature review. J Vocat Educ Train 70(3):337–363

Bächle TC, Bareis J (2022) "Autonomous weapons" as a geopolitical signifier in a national power play: analysing AI imaginaries in Chinese and US military policies. Eur J Futures Res 10(1):20. https://doi.org/10.1186/s40309-022-00202-w

Balbi G, Leggero R (2020) Communication is maintenance: turning the agenda of media and communication studies upside down. H-Ermes J Commun 17:7–26

Barassi V, Scharenberg A, Poux-Berthe M, Patra R, Di Salvo P (2022) The human error project. Research report AI errors and the profiling of humans: mapping the debate in European news media. Available at https://thehumanerrorproject.ch/wp-content/uploads/2022/12/Research-Report_AI-Errors-Mainstream-Media_Human-Error-Project.pdf. Accessed 24 April 2024

Bareis J, Katzenbach C (2021) Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. Sci Technol Human Values 47(5):855–881

Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: Can language models be too big?. In: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, pp 610–623

Bory P (2019) Deep new: The shifting narratives of artificial intelligence from Deep Blue to AlphaGo. Convergence: Int J Res New Media Technol 25(4):627–642. https://doi.org/10.1177/1354856519829679

Bory P (2020) The internet Myth: From the Internet Imaginary to Network Ideologies. University of Westminster Press, London

Bory P, Natale S, Trudel D (2021) Artificial intelligence: Reframing thinking machines within the history of media and communication. In: Balbi G, Ribeiro N, Schafer V, Schwarzenegger C (Eds.) Digital Roots. De Gruyter Oldenbourg, Berlin, pp 95–114. https://doi.org/10.1515/9783110740202

Bostrom N (2014) Superintelligence. Oxford University Press, Oxford

Broussard M (2018) Artificial unintelligence: how computers misunderstand the world. MIT Press, Cambridge (MA)

Bucher T (2017) The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. Inf Commun Soc 20(1):30–44

Bunz M, Braghieri M (2022) The AI doctor will see you now: assessing the framing of AI in news coverage. AI Soc 37(1):9–22

Capraro V et al. (2024) The impact of generative artificial intelligence on socioeconomic inequalities and policy making. PNAS Nexus (forthcoming).

Castoriadis C (1997) The imaginary institution of society. MIT Press, Cambridge (MA)

Cave S, Dihal K, Dillon S (eds) (2020) AI narratives: a history of imaginative thinking about intelligent machines. Oxford University Press, Oxford

Center for AI Safety (2023) Statement on AI Risk. May 30. https://www.safe.ai/work/statement-on-ai-risk. Accessed 24 April 2024

Chinese Aerospace Studies Institute (2017) A next generation artificial intelligence development plan. https://www.airuniversity.af.edu/Portals/10/CASI/documents/Translations/2021-03-02%20China%27s%20New%20Generation%20Artificial%20Intelligence%20Development%20Plan-%202017.pdf. Accessed 24 April 2024

Chubb J, Reed D, Cowling P (2022) Expert views about missing AI narratives: is there an AI story crisis? AI Soc. https://doi.org/10.1007/s00146-022-01548-2

Cools H, Van Gorp B, Opgenhaffen M (2024) Where exactly between utopia and dystopia? A framing analysis of AI and automation in US newspapers. Journalism 25(1):3–21

Crary A (2023) The toxic ideology of longtermism. Radic Philos 2(14):49–57

Crawford S (2007) Internet think. J Telecommun High Technol Law 5:467–486

Crawford K (2021) The Atlas of AI: power, politics, and the planetary costs of artificial intelligence. Yale University Press, New Haven

Crevier D (1993) AI: the tumultuous history of the search for artificial intelligence. Basic Books Inc, London

Cugurullo F (2024) The obscure politics of artificial intelligence: a Marxian socio-technical critique of the AI alignment problem thesis. AI Ethics. https://doi.org/10.1007/s43681-024-00476-9

Depounti I, Saukko P, Natale S (2022) Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend. Media Cult Soc 45(4):720–736. https://doi.org/10.1177/01634437221119021

Di Summa L (2019) Black mirror: the not so fearful consequences of technology. Film Philos 23:95–113

Dreyfus HL (1972) What computers can't do: a critique of artificial reason. Harper & Row, New York

Ernst C (2019) Beyond meaningful human control? – Interfaces und die imagination menschlicher kontrolle in der zeitgenössischen Diskussion um autonome Waffensysteme (AWS). In: Thimm C, Bächle TC (eds) Die Maschine: Freund oder Feind? Mensch und Technologie im digitalen Zeitalter. Berlin, Springer Fachmedien, pp 261–299. https://doi.org/10.1007/978-3-658-22954-2_12

European Union (2024) 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. Interinstitutional File: 2021/0106(COD). Brussels, 26 January 2024.

Flichy P (2007) The internet imaginaire. MIT press, Cambridge (MA)

Future of Life Institute (2023) Pause Giant AI experiments: an open letter. https://futureoflife.org/open-letter/pause-giant-ai-experiments/. Accessed 24 April 2024

Galanos V (2023) To have done with the metaphor of summers and winters: can AI and internet history cure hype? SSRN 4640305. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4640305. Accessed 24 April 2024

Garcia JL (ed) (2014) Pierre musso and the network society: from saint-simonianism to the internet. Springer, Champaign-Urbana (IL)

Ge H, Wu Y (2023) An empirical study of adoption of ChatGPT for bug fixing among professional developers. Innov Technol Adv 1(1):21–29

Gebru T, Mitchell M (2022) We warned Google that people might believe AI was sentient. Now it's happening. Washington Post, 17 June. https://www.washingtonpost.com/opinions/2022/06/17/google-ai-ethics-sentient-lemoine-warning/. Accessed 24 April 2024

Gebru T, Bender EM, McMillan-Major A, Mitchell M (2023) Statement from the listed authors of Stochastic Parrots on the "AI pause" letter, 31 March 2023. https://www.dair-institute.org/blog/letter-statement-March2023/. Accessed 24 April 2024

Gebru T, Torres EP (2024) The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. First Monday. https://doi.org/10.5210/fm.v29i4.13636

German Parliamentary Expert Commission (2021) Artificial intelligence – social responsibility and economic, social and ecological potential. Report of the German parliamentary expert commission. Deutscher Bundestag Drucksache 19/23700, 28 October.

Goode L (2018) Life, but not as we know it: AI and the popular imagination. Culture Unbound: J Curr Cult Res 10(2):185–207

Granovetter MS (1973) The strength of weak ties. Am J Sociol 78(6):1360–1380

Greengard S (2023) Computational linguistics finds its voice. Commun ACM 66(2):18–20

Gunkel DJ (2012) The machine question: critical perspectives on AI, robots, and ethics. Mit Press, Cambridge (MA)

Jasanoff S, Kim SH (2015) Dreamscapes of modernity: sociotechnical imaginaries and the fabrication of power. University of Chicago Press, Chicago

Jones B, Jones R, Luger E (2022) AI 'Everywhere and Nowhere': addressing the AI intelligibility problem in public service journalism. Digit J 10(10):1731–1755

Kurzweil R (2005) The singularity is near: when humans transcend biology. Penguin, London

Lemoine (2022) Is LaMDA sentient? — an interview. Available at https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917. Accessed 24 April 2024

Licklider JC (1960) Man-computer symbiosis. IRE Trans Hum Fact Electron 1:4–11

Lomborg S, Kapsch PH (2019) Decoding algorithms. Media Cult Soc 42(5):745–761

Mager A, Katzenbach C (2021) Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. New Media Soc 23(2):223–236. https://doi.org/10.1177/1461444820929321

Mansell R (2012) Imagining the internet: communication, innovation, and governance. Oxford University Press, Oxford

Marcus GE (ed) (1995) Technoscientific imaginaries. University of Chicago Press, Chicago

Martin CD (1993) The myth of the awesome thinking machine. Commun ACM 36(4):120–133

Marx L (1964) The machine in the garden: technology and the pastoral ideal in America. Oxford University Press, New York

Minsky M (1961) Steps toward artificial intelligence. Proc IRE 49(1):8–30

Mosco V (2005) The digital sublime: myth, power, and cyberspace. Mit Press, Cambridge (MA)

Narayanan A (2019) How to recognize AI snake oil. Arthur Miller lecture on science and ethics.

Natale S (2021) Deceitful media: Artificial intelligence and social life after the Turing test. Oxford University Press, Oxford

Natale S, Ballatore A (2020) Imagining the thinking machine: Technological myths and the rise of artificial intelligence. Converg 26(1):3–18

Natale S, Cooke H (2020) Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces. Media Cult Soc 43(6):1000–1016. https://doi.org/10.1177/0163443720983295

Nye DE (1996) American technological sublime. Mit Press, Cambridge (MA)

Paltieli G (2022) The political imaginary of national AI strategies. AI & Soc 37(4):1613–1624

Parviainen J, Coeckelbergh M (2020) The political choreography of the Sophia robot: beyond robot rights and citizenship to political performances for the social robotics market. AI & Soc. https://doi.org/10.1007/s00146-020-01104-w

Prillaman M (2024) Is ChatGPT making scientists hyper-productive? The highs and lows of using AI. Nature 627:16–17. https://doi.org/10.1038/d41586-024-00592-w

Rasskin-Gutman D (2009) Chess metaphors: artificial intelligence and the human mind. MIT Press, Cambridge (MA)

Rhee J (2018) The robotic imaginary: the human & the price of dehumanized labor. University of Minnesota Press, Chicago

Rip A, Voß JP (2013) Umbrella terms as mediators in the governance of emerging science and technology. Sci Technol Innov Stud 9(2):40–59.

Russell SJ (2010) Artificial intelligence a modern approach. Pearson Education, London

Santoro E, Monin B (2023) The AI effect: people rate distinctively human attributes as more essential to being human after learning about artificial intelligence advances. J Exp Soc Psychol 107:104464. https://doi.org/10.1016/j.jesp.2023.104464

Sartor G, Omicini A (2016) The autonomy of technological systems and responsibilities for their use. In: Kreβ C, Liu H-Y, Bhuta N, Geiβ R, & Beck S (Eds.) Autonomous weapons systems: law, ethics, policy. Cambridge University Press, Cambridge, pp 39–74. https://doi.org/10.1017/CBO9781316597873.003

Schulz C (2023) A new algorithmic imaginary. Media Cult Soc 45(3):646–655

Searle JR (1980) Minds, brains, and programs. Behav Brain Sci 3(3):417–424

Skjuve M, Følstad A, Fostervold KI, Brandtzaeg PB (2021) My chatbot companion: a study of human-chatbot relationships. Int J Hum Comput Stud 149:102601

Sobchack VC (1997) Screening space: the American science fiction film. Rutgers University Press, New Brunswick

Sormani P (2023) Interfacing AlphaGo: embodied play, object agency, and algorithmic drama. Soc Stud Sci 53(5):686–711

Sparkes M (2022) No sign of a machine mind yet. New Sci 254(3391):9. https://doi.org/10.1016/S0262-4079(22)01039-9

Star SL (1999) The ethnography of infrastructure. Am Behav Sci 43(3):377–391

Streeter T (2011) The net effect: Romanticism, Capitalism, and the internet. NYU Press, New York

Suchman L (2023) The uncontroversial 'thingness' of AI. Big Data Soc. https://doi.org/10.1177/20539517231206794

Taylor C (2004) Modern social imaginaries. Duke University Press, Durham (NC)

Tiku N (2022) The Google engineer who thinks the company's AI has come to life. *The Washington Post* 6 June 2022. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/. Accessed 24 April 2024

Turing AM (1950) Computing machinery and intelligence. The essential turing: the ideas that gave birth to the computer age (2012):433-464

UK Government (2021) National AI strategy. London, September 2021,. https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf. Accessed 24 April 2024

White House Office of Science and Technology Policy (2018) Summary of the 2018 White House Summit on artificial intelligence for American industry. https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf. Accessed 24 April 2024

Wiener N (1966) God & Golem Inc: a comment on certain points where cybernetics impinges on religion. MIT press, Cambridge (MA)

Woods HS (2018) Asking more of Siri and Alexa: feminine persona in service of surveillance capitalism. Crit Stud Media Commun 35(4):334–349. https://doi.org/10.1080/15295036.2018.1488082

Ytre-Arne B, Moe H (2021) Folk theories of algorithms: Understanding digital irritation. Media Cult Soc 43(5):807–824. https://doi.org/10.1177/0163443720972314