RESEARCH ARTICLE

Synthetic Socio-Technical Systems: Poiêsis as Meaning Making

Piercosma Bisconti[1,5] · Andrew McIntyre[2] · Federica Russo[3,4]

Abstract

With the recent renewed interest in AI, the field has made substantial advancements, particularly in generative systems. Increased computational power and the availability of very large datasets has enabled systems such as ChatGPT to effectively replicate aspects of human social interactions, such as verbal communication, thus bringing about profound changes in society. In this paper, we explain that the arrival of generative AI systems marks a shift from 'interacting through' to 'interacting with' technologies and calls for a reconceptualization of socio-technical systems as we currently understand them. We dub this new generation of socio-technical systems *synthetic* to signal the increased interactions between human and artificial agents, and, in the footsteps of philosophers of information, we cash out agency in terms of 'poiêsis'. We close the paper with a discussion of the potential policy implications of synthetic socio-technical system.

Keywords Generative AI · Socio-technical systems · Poietic agent · Philosophy of technology · Actor-network theory

✉ Federica Russo
f.russo@uu.nl

Piercosma Bisconti
piercosma.bisconti@dexai.eu

Andrew McIntyre
a.mcintyre@uva.nl

[1] Artificial Intelligence and Intelligent Systems Lab, Interuniversity Consortium for Computer Science, Rome, Italy

[2] University of Amsterdam, Amsterdam, The Netherlands

[3] University of Utrecht, Utrecht, The Netherlands

[4] University College London, London, UK

[5] DEXAI - Artificial Ethics, Rome, Italy

🌀 Springer

1 Introduction

Recently, artificial intelligence (AI) systems have made significant and unprecedented advances that could not have been predicted even a few years ago, and now this technology is taking on a more substantial role in society with skills and behaviours that increasingly resemble those of human agents. For instance, where computational humour was a nascent avenue of study within computer science only a few years ago (Hempelmann et al., 2006; Stock & Strapparava, 2003), contemporary large language models (LLMs) such as OpenAI's ChatGPT have now become adept at identifying recurring patterns of humour and can replicate these patterns to generate original jokes (Chowdhery et al., 2022). Beyond computational humour, there are numerous other examples of AI-generated content (e.g., automated poetry, generative portraiture) that demonstrate how these technologies have become proficient at replicating aspects of human social interactions that are intricate, context-specific and often steeped in cultural symbolism. In contrast to a few decades ago, when information communication technologies were primarily limited to specific contexts or purposes, the versatility and interactivity of today's AI technologies has led to the development of general-purpose programs that can be used in a diverse range of situations and tasks such that these technologies are rapidly becoming deeply integrated within our social environment.

With this deep integration comes a slew of challenges for Western democracies as, for example, AI systems can be used to bolster disinformation campaigns contributing to the climate of uncertainty and societal division of post-truth politics (Coeckelbergh, 2023; Rini, 2020). Even when used in good faith, these programs are unpredictable and often return bias or inaccurate content thus perpetuating discriminatory ideologies (Bommasani et al., 2022; Chun, 2021; Crawford, 2021). The unpredictability of the outputs has prompted policymakers around the world to turn their attention toward foundational models and general-purpose AI systems, particularly those used in the generation of text, images and audiovisual content. However, the societal impact of integrating AI systems into societies is unclear and has proven difficult to properly evaluate for the purposes of policymaking, due in part to AI's inherent unpredictability. Thus, researchers have come to consider theoretical frameworks able to capture the interplay between technology and societies, such as the concept of socio-technical systems (Ropohl, 1999), an approach that recognizes the reciprocal relationship between technology and society.

Building on these theoretical suggestions, we argue in this paper that assessment of the unique societal impact of AI-generated content requires a new theoretical understanding of how generative AI interact with other actors within a social system. We propose to relabel socio-technical systems involving generative AI as *synthetic* to signal that human and artificial agents are not merely equal and legitimate actors within societies (an idea that we borrow from Actor-Network Theory) but also *interact* in unprecedented ways. From the perspective of human agents, the arrival of generative AI marks a shift from technologies that we interact *through* to technologies we interact *with*, in such a way that goes beyond the idea of quasi-otherness proposed by Ihde (1990). To fully explain the significance of this shift and the relevance of the 'synthetic' label, we mobilise the concept of 'poietic agent' that we borrow from the

philosophy of information. The main claim advanced in this paper is that current and future generative AI systems are capable of creating new semantic artifacts that contribute to our collective knowledge and circulate in social systems in the same way as those created by humans, thus marking a qualitative shift from previous technologies that were only capable of mediating between human actors in social system.

To begin, Sect. 2 discusses how the growing integration of AI technology in society is re-shaping our understanding of socio-technical systems. As existing theoretical frameworks are inappropriate for analysing the complex challenges of these new socio-technical systems, we argue for a system-level approach that builds upon the framework of Actor-Network Theory (ANT), which we describe in Sect. 3. ANT paves the way for a fundamental rethink of the social interactions between human and non-human actors, but it remains limited in its application. Thus, Sect. 4 introduces and discusses the notion of the 'poietic agent' as a conceptual tool for describing the mechanisms underlying those socio-technical systems that, when merged with the ANT framework, allows for a deeper understanding of AI as social actors, and notably their role in *synthetic* socio-technical systems. In Sect. 4, we put forth the main differences between generative AI and other technologies in shaping socio-technical systems. Section 5 concludes with a discussion on how this approach may begin to shape AI policymaking.

2 AI Systems as Social Actors

Where traditional approaches in philosophy of technology focus on the impact of a specific technology on a specific context, socio-technical systems theory proposes a multi-directional analytical perspective that does not consider technology and social systems as isolated entities separated from one another (Ropohl, 1999; Trist, 1981). Rather, the fabric of a social system guides the development and design of technologies and, in turn, those technologies impact upon and re-shape those same social systems from which they emerged. This theoretical approach is especially pertinent when analysing the impact of AI technologies, given how broadly and deeply this technology has begun to permeate human society. The widespread deployment of AI systems across various social domains presents an immense challenge for analysing impact. These technologies have now become so pervasive in society, often interacting with one another (e.g., AI-driven stock trading), that they cannot be analytically isolated within a specific context or social domain. For example, chatbots capable of producing coherent text (e.g., ChatGPT) are simultaneously being used for customer relations in private business, as writing tools in journalism and academia, and as interactive instructors in education. Meanwhile, text-to-image programs (e.g., Midjourney) are not restricted to the artworld but are further used in advertising, data visualization, and pornography. Furthermore, pattern recognition programs have a wide variety of uses including predicting market trends, surveillance, scientific research and medical diagnosis (e.g., identifying rare skin cancers). Additionally, AI-driven social robotics has recently seen similar advancements (e.g., intelligent physical interactions, autonomous motivation, adept navigation within dynamic environments, and dexterous object manipulation) that will allow these technologies

to be integrated into a diverse range of social settings from hospitality to the care sector. With the pervasive integration of AI systems comes an increase in the complexity of our socio-technical systems that makes determining the societal impact of a specific technology or a particular AI program a Quixotic endeavour. However, it is not simply the complexity of the socio-technical system that presents a challenge. With an increased capacity to navigate the interactional and functional nuances of human social systems autonomously and adaptively, AI is becoming an *actor* with characteristics that are very similar to human actors, and therefore legitimately qualifies as *social*.

The emergence of AI as a social actor represents a significant shift from previous socio-technical systems. Historically, information and communication technologies were primarily seen as tools through which humans interacted with one another. With AI, however, humans are able to directly interact with the technology itself. Previously, the role of these technologies was limited to either reliably transmitting or distorting messages.

As philosopher and sociologist of science Bruno Latour highlights, the social world has been mediated by non-human actors (Latour, 1996, 2005) (this notion will be discussed in greater detail later) since the Palaeolithic Age but these actors remained in the background of sociality, being merely the facilitators or inhibitors of social experience. With the emergence of sophisticated AI systems, however, the social world that once pertained solely to humans, and perhaps their anthropomorphized pets (Wang et al., 2022), is rapidly changing. AI social actors, be they embodied in a physical form (e.g., social robots (Duffy et al., 1999) or lacking embodiment entirely (e.g., deepfakes (Westerlund, 2019), chatbots), are moving to the foreground of sociality and are no longer objects that humans interact *through* but social actors they interact *with*. While we might commonly state that one writes with a pen or paints with a brush, these objects are not social actors in themselves but rather humans interact with one another through these technologies. In stark contrast, AI social actors are increasingly recognized as fully developed participants that humans interact *with* in a whole range of social interactions, not as mere message deliverers or facilitators of communication. Non-human actors have thus taken on an entirely new role within social systems since they can contribute to building narratives, alter relationships between facts and their interpretations directly, shape social values by discussing them, and contribute to the broader ecosystem of interactions that constitute our social systems – all of this, as we shall discussion in Sect. 4, has to do with *meaning making*, encapsulated in the concept of poiêsis. Differently put, we have begun to properly interact *with* these artificial agents *as if* they were human agents (Seibt, 2017), while still recognizing their artificiality.

As more and more AI systems capable of mimicking human social behaviour become integrated into society, it is not the case that they will simply replace those current human social actors and fulfil the same role and interactions. AI social actors are not equivalent to human actors but may be understood as *quasi-others*, a concept developed by philosopher of technology Don Idhe and others in recent years (Coeckelbergh, 2011; Gunkel, 2017; Ihde, 1990): a quasi-other is an agent that can be perceived as a human-like individual. Certainly, not all AI technologies trigger perceptions of quasi-otherness (e.g., a medical decision support system is not viewed

as a social counterpart) but, for the purposes of this discussion, in this paper we only focus on those that do fit this mould. Due to the explicitly social capabilities of AI discussed above, the interactions between humans and AI systems cannot be described as being purely instrumental (interact *through*) in the same manner as other human-technology relationships that characterize the technology as a simple tool to be used, such as a person driving a car or using a laptop. However, neither can these interactions be understood as fully intersubjective in the same manner as human-human relationships. On the one hand, this is due to technical limitations (e.g., process lag, system failures) preventing smooth and seamless interactions with programs. On the other hand, often this is due to one's own conscious awareness that they are interacting with a non-human, as might be observed in relation to the uncanny valley effect (Kätsyri et al., 2015; Mori, 1970; Walters et al., 2008) and other instances in which humans have been unwilling to interact with social robots and artificial social actors as they would with other humans (Naneva et al., 2020; Nomura et al., 2008). In fact, humans are usually prevented from interacting with AI systems in an intersubjective way because they are perceived as having no intentionality, a crucial aspect of human-human interactions. Rather, the human-AI relationship lies somewhere between these two poles with the AI perceived to be neither object nor other but a quasi-other that allows for something that closely resembles an inter-subjective relationship (Bisconti & Carnevale, 2022; Coeckelbergh, 2016; Gunkel, 2017). As such, the pervasive integration of quasi-other technologies as social actors within our social systems will not replace human actors but will alter the nature of the interactions and dynamics of the social system itself.

To account for this fundamental change, we propose to define these new socio-technical systems as *synthetic* in that non-human social actors are increasingly being integrated into every aspect of the system and, therefore, they can no longer be recognized as purely human social systems. In fact, nowadays, at least in technologically developed societies and increasingly all over the world, a fundamental part of human social interactions happen through digital intermediators, be they social networks, telephones or video conferencing software, among others. At the same time, the digital space is increasingly populated by AI-driven actors, that both interact with us (e.g. chatbots), allow our interactions in the digital space (e.g. AI-driven traffic management of 5G infrastructure), and produce content that we enjoy (e.g. GPT texts, DALL-E images). It is clear from these developments that non-human social actors are progressively altering every aspect of human social systems such that they may be considered *synthetic*. However, the fundamental aspect we want to focus on is the following: where previous non-human actors were largely predictable in their behaviour and interactions with their environment, AI systems introduce a level of unpredictability reminiscent of human behaviour due to their adaptability, autonomy and growing prominence in social interactions. On multiple occasions in the following discussion, we refer to the concepts of agency, autonomy and adaptability as characteristics distinguishing AI systems from 'old-fashioned artifacts', so let us clarify these terms. When referring to the agency of AI systems, we follow the meaning provided in the Actor-Network Theory context introduced in Sect. 3. Briefly put, agency is understood only as the ability of an actor to modify the system of relations between other actors, a meaning that is close to that given within the philosophy

of information, relying on interactivity, and that does not presuppose or entail consciousness or intentionality. For what concerns autonomy, this is understood both as the ability of AI systems to perform tasks without human intervention (Ezenkwu & Starkey, 2019), and the fact that the relationship between inputs and outputs are not fully predictable. Adaptability, very much connected with the previous two, is understood as the ability to change behaviour to maintain performance under environmental changes. Adaptability in AI usually implies the ability to learn new patterns and relationships in the outer world.

With these novel characteristics come urgent real-world challenges such as the potential for AI-generated disinformation to be used to manipulate public sentiment, interfere in democratic elections, and incite violence or political strife (Chesney & Citron, 2019; Diakopoulos & Johnson, 2021). Thus, there is a pressing need to develop strategies and regulations to manage the influence of AI social actors and to guarantee safety and trustworthiness. However, existing analytical frameworks for analysing human-machine relationships appear inappropriate for considering synthetic socio-technical systems. Verbeek (2016), among others, proposed post-phenomenological approaches to the human-machine relationship that characterize technologies as *mediators* between humans and the world around them. While an advantage of this approach is that it avoids the anthropocentric notion of technologies as tools or instruments, it seems insufficient when confronted with AI social actors that are, as discussed above, things that humans interact *with* rather than *through*.

Approaches from philosophy of information, particularly the work of Floridi (2014), may seem more suitable in that they analyse technologies in terms of their *in-betweenness*, a property that describes how a given technology relates a human agent to an environment, via a prompt. For example, harsh sunlight is the prompt of the in-between technology of sunglasses in that it invites the person to wear them; this is called, in the terminology introduced by Floridi, a first-order technology and the relations between this technology, humans and nature take the following form:

$$humans \leftarrow technology \rightarrow nature$$

Most first order technologies are analogue technologies that humans have designed to cope in a natural environment. *Second*-order technologies are instead more sophisticated in that a technology is 'in between' humans and another technology:

$$humans \leftarrow technology \rightarrow technology$$

For instance, a television is the prompt of the in-between technology of a remote control in that it invites the user to press the buttons. However, with AI systems, and in general with digital tools with high degrees of agency and autonomy, human agents may be seen as external to these sophisticated technologies, as both prompts and agents are technologies themselves, which Floridi describes as a third-order relationship. An example would be those AI stock exchange programs that make decisions to sell and buy stocks autonomously and without human intervention. Therefore, in Floridi's account, human agents do interact with technologies, but when we reach

third-order technologies, it may seem, in their schematic representation, that humans are out of the loop entirely:

$$technology \leftarrow technology \rightarrow technology$$

This perspective, however, conveys the idea that humans have a role as users of AI systems, but still may be well outside this chain. We need instead to take a broader look at third-order interactions and explicitly acknowledge that humans play significant roles in the production and usage of these technologies in society including designers, developers, and regulators, as our argument in the next section will highlight. In one way or another, humans are always in the loop (Russo, 2022). As these existing frameworks show limitations with respect to the new challenges that synthetic socio-technical systems pose, it is necessary to formulate a new system-level framework that acknowledges the role of AI as social actors but that further understands these programs as technologies that are always situated within human social environments. This system-level framework has the objective to clarify: (i) the extent to which AI systems and 'old-fashioned artifacts' differ in bringing modifications to social systems, and (ii) if they do differ, which are the peculiar characteristics of AI systems and how they will impact and modify social systems.

To move toward such a framework, we build on Actor Network Theory (Latour, 2005), that we introduce next.

3 The ANT Basis of Synthetic Socio-Technical Systems

Closely associated with the work of Bruno Latour (1992, 1996, 2002, 2005), ANT offers a radical perspective on the forces that shape our social realities. At its core, ANT seeks to dissolve the traditional boundaries of sociology, emphasizing the profound interconnectedness of all social entities within a given network. In the terminology of ANT, 'actors' or 'actants' are not solely identified as humans but further refer to a broad spectrum of entities, including people, artefacts, technologies, and institutions, that interact with one another in a flat non-hierarchical network. This inclusive approach underscores the foundational belief of ANT that every actor, be they human or non-human, has the capacity to act, to influence, and to instigate change in the network as a whole. Thus, ANT is of particular interest to our enquiry into synthetic socio-technical systems as it dispenses with the traditional, anthropocentric view of the human-machine relationships in which AI would be understood as a simple tool to be intentionally used by humans and allows for a more nuanced description of the explicitly social function of these technologies. In this paper, we take 'actors' and 'agents' as synonymous, but overall, we prefer the term 'agent' to emphasise actors' abilities to modify a system through their interactions. It is also worth noting that no actor exists in a vacuum; therefore, every actor is, by definition, a social actor. The term 'social actor' is used specifically to emphasize that this actor possesses interactional abilities that allow it to modify systems in ways that are similar to human capabilities, such as producing meaningful semantic content, which is our focus in this paper.

To illustrate the application of Actor-Network Theory (ANT), we analyze the social system involved in the creation, distribution, and reception of AI-generated content online, as depicted in Figs. 1 and 2. At the moment of engaging with this content, a variety of social actors—both human and non-human—converge and interact within a network. This network, characterized by its complexity and the multitude of actors involved, is connected through intricate relationships. The diagrams in Figs. 1 and 2 offer a simplified view to identify key social actors, describe their distinct characteristics, and demonstrate how they are interconnected within the network. Figure 1 provides a basic overview of the different groupings of social actors and how these are typically understood to interact with each other. These include social actors involved in the development and distribution of a generative AI program, the creation of AI-generated content using this program, the training data to be replicated, the circulation of this content in online spaces, the user reception of the content, the various policy and legislative interventions, and the broader public discourse surrounding the AI-generated content. Each of these groupings can be further unpacked to reveal a multitude of social actors at play, as shown in Fig. 2.

The arrows shown in Figs. 1 and 2 indicate a possible pipeline involved in the production, circulation, and reception of AI-generated content, with the motivations and characteristics of each social actor impacting upon the next in the sequence. To begin, there are different kinds of generative AI developer (e.g., private companies, academic institutions, individual programmers, government bodies) and each displays unique characteristics and motivations that define their role within the broader network. These characteristics include the developer's motivations for producing the program (e.g., profitable product, public service, industry innovation), their access to resources (e.g., funding, expertise), and their adherence to particular industry standards and regulations (e.g., AI for Social Good principles). Particularly, these characteristics will directly impact upon the technical design of the program itself. For example, an academic institution may be motivated to create a program for technical experts to use in research and thus seek to produce a high-quality product, but one
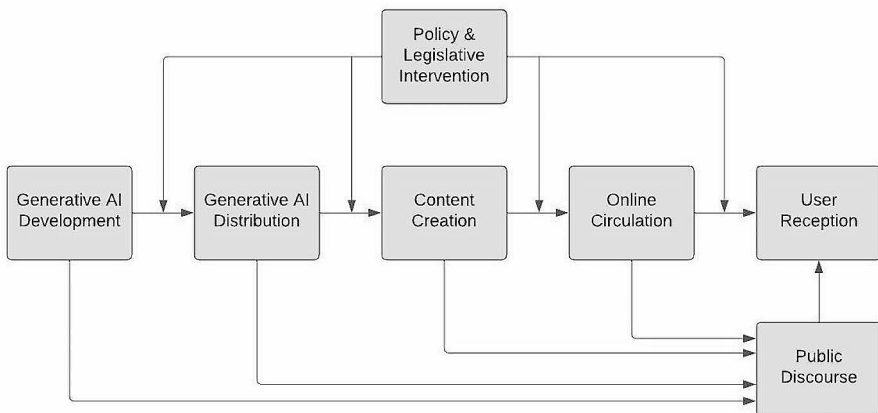

Fig. 1 A general approximation of the significant groups of social actors involved in the production, circulation and reception of AI-generated content online
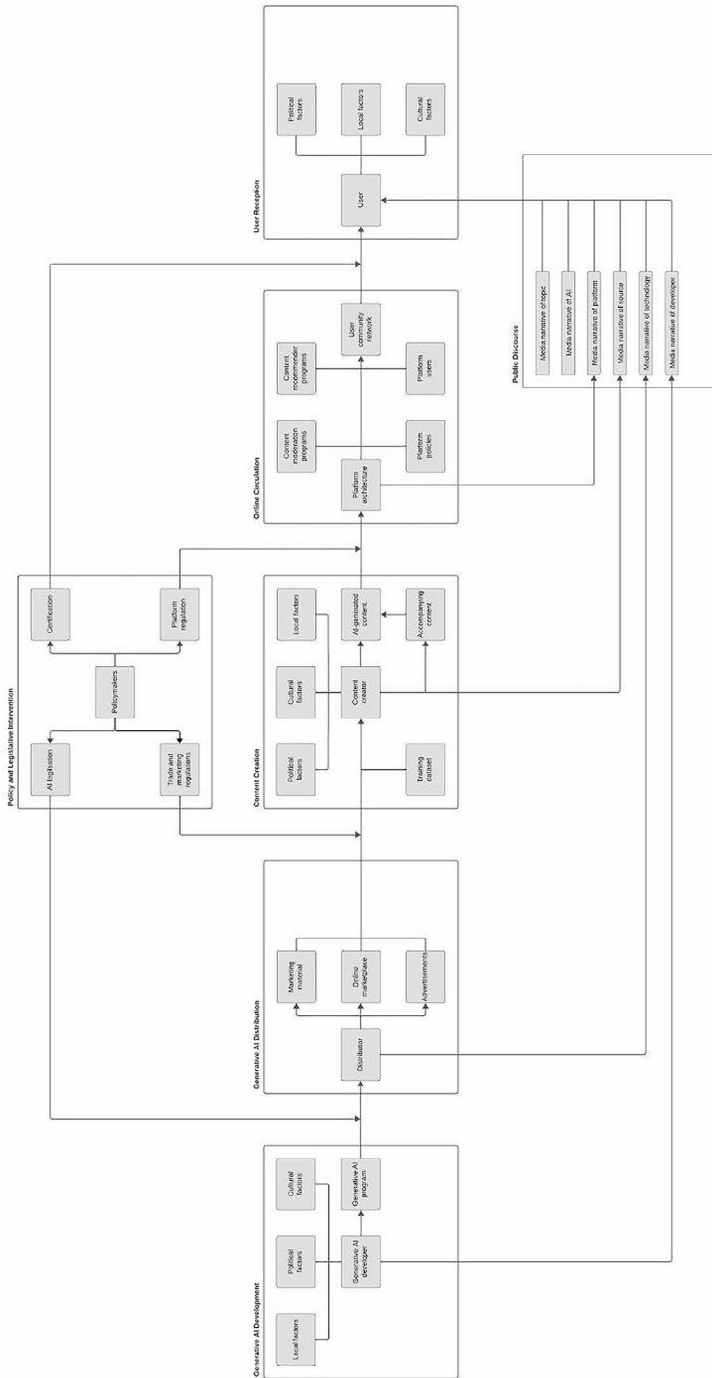
Fig. 2 An expanded view of the network of social actors involved in the production, circulation and reception of AI-generated content online

that is perhaps complicated or cumbersome for non-experts to use. In contrast, a private company may be motivated to create a program to maximise profits and thus seek to produce a user-friendly but low-quality product. These specific capabilities of the program itself further influence how it is marketed and distributed to the public in that it frames how the technology is intended to be used, for example a user-friendly generative AI app may be advertised as an administrative tool for businesses by showcasing how a user can easily use the app to generate emails and reports with basic written prompts. Of course, it is entirely up to those content creators using the technology to determine the substance of the AI-generated content itself by providing prompts to the program as well as the paratextual elements (e.g., titles, links, descriptions) accompanying the content when published online. However, the quality of the content is, to some extent, limited by the size and quality of the available training data.

Once published online, how AI-generated content circulates on social media platforms and among users is determined by the specific architecture of the platforms themselves and the different information channels available to users (e.g., posts, comments sections, private messaging), as well as those automated programs moderating and recommending content according to platform policies. It is via these technical structures that the AI-generated content eventually reaches a user's computer screen but how they perceive it (e.g., whether or not they deem it true) is further influenced by the reactions (e.g., likes, dislikes) and comments of other platform users, particularly those in the user's immediate online network (e.g., friends, family, colleagues). Finally, at every stage of this process there are possible policy and legislative interventions such as specific AI legislation (e.g., EU AI Act) that may interfere with the diffusion of AI-generated content. Each of the above actors is influenced by complex political, cultural and local factors in their social environment but the exact nature of their role within synthetic socio-technical systems is beyond the scope of this paper.

The perception and belief in AI-generated content is influenced by the characteristics of the user, such as demographics and media literacy. Additionally, these perceptions may be shaped by broader public discourse and prevailing narratives found in other media, such as news stories, films, advertisements, and interviews. There are, of course, numerous media narratives to consider but those of particular interest in this case relate to the generative AI developer (e.g., brand reputation), the product itself (e.g., marketed usage), the content creator (e.g., trustworthiness, political affiliation), the platform on which the content is circulated (e.g., safety standards), the topic addressed in the AI-generated content (e.g., controversy), and AI technology more generally (e.g., capabilities). Primarily, these narratives work to establish an actor's reputation or expectations about that actor that will influence the reception of the AI-generated content. The impact of these media narratives is beyond the scope of this paper but it is necessary to note that the content produced by generative AI does not exist in isolation but rather it exists as a part of a broader information environment in which it interacts with other content in complex ways that alter the knowledge it contributes to the socio-technical system and how it is understood by human agents.

Until this point, this section has only described and discussed the progression of activities but, when considered through the lens of ANT, networks are far from static structures. As mentioned, ANT maintains that non-human social actors, such as those

technologies identified above (e.g., generative AI systems, platform architectures, AI-generated content), are not simply passive tools used by human developers and content creators but are *active* participants equally capable of influencing the network. Networks are dynamic, ever-evolving constellations of relationships, defined and redefined by the ceaseless interactions of their constituent actors. The very essence of a network is shaped by the continuous negotiations and interactions between these actors. Within these networks a series of negotiations and transformations unfold that determine the roles and relationships of the actors involved. Furthermore, Latour's concept of 'relational materiality' suggests that the identities and characteristics of actors are not fixed or inherent. Instead, they emerge from their relationships with other actors. In this view, identity becomes fluid, shaped by one's position and connections within the broader network. This is where Figs. 1 and 2 are limited in that the lines and arrows connecting these social actors are continually redrawn and multi-directional depending on the context.

This fluidity in relationships presents a notable challenge for using ANT to analyse synthetic socio-technical systems as, while it enables us to adeptly outline which actors interact within a network, it fails to detail the exact nature of those interactions. Furthermore, the non-hierarchical approach of ANT presents human and non-human actors as being equally capable of shaping the network in which they exist. While the influence of non-human actors should be properly accounted for, this apparent 'flattening' of synthetic socio-technical systems means that it is difficult to distinguish between human social actors and quasi-other technologies such as generative AI. ANT allows for an initial description of a synthetic socio-technical system, as provided above, but does not elucidate the intricate nature of the relationships between actors, and the normative considerations that may stem from it. Without greater insight into the internal dynamics of the interactions between these actors, we can only observe the outcomes of synthetic socio-technical systems without grasping the causal processes that brought about these outcomes. ANT lacks the necessary depth for practical application and informed policymaking as we remain mere observers noting occurrences without fully understanding the reasons behind them. To advance our analysis beyond ANT, in the next section we introduce the concept of 'poiêsis', which is able to provide greater insight into how actors interact, and will specify the peculiar way, compared to old-fashioned artifacts, in which AI systems provoke system-level modification.

4 Poietic Agents as Conceptual Tools

In ANT, actors can either mediate or intermediate social systems. Those that mediate the social system modify the network of relationships, while those that intermediate leave it unchanged. This distinction raises two pivotal questions. Firstly, when an actor originally happens to either mediate or intermediate, which systemic elements drive this difference, and what underlying processes determine this selection? Secondly, when considering how a network is either constructed or deconstructed through mediation, what mechanisms govern these network dynamics and what specific changes in information content qualify as mediation and why? While the ques-

tions are important, the answers do not reside in ANT theory per se. In this section, we sketch the contours of a theoretical approach that helps shed light on these questions.

Building on Floridi's notion of the *homo poieticus* (Floridi, 2013; Floridi & Sanders, 2005) and integrating the work of Simondon (2020), Russo (2022) introduces the concept of the "poietic agent" to better conceptualise the way in which human and artificial agents produce knowledge together through a collaborative process. The concept of 'poietic agent' will help us understand, from a theoretical perspective, how AI systems differ from traditional technologies in influencing social systems, and thus allow us to delve deeper into the dynamics of synthetic socio-technical systems and their functionality.

To elucidate the partial autonomy and consequent 'agentiality' of technical objects, Russo engages with the works of both Latour and Simondon. From Latour she borrows a basic ANT approach, as we do in Sect. 3, while from Simondon she borrows the idea that artificial actors capable of agency are not just digital technologies, but also analogue ones. Our arguments, however, primarily concern digital technologies. Of particular interest to us are the poietic characteristics that Russo ascribes to artificial agents. Developing on arguments previously established in philosophy of information (Floridi, 2013), Russo argues that as artificial agents are able to process information, they qualify as epistemic agents which are capable of altering the dynamics between other agents and their environment. In the context of our discussion on synthetic socio-technical systems, we are interested in the ability of artificial agents to alter the communication environment through the generation of text and audiovisual content and, more generally, through the generation of semantic content.

Russo's perspective on epistemic agency is inherently relational and influences how one should conceive of 'knowledge'. In turn, her approach to knowledge is rooted in Floridi's *constructionism* (Floridi, 2011), a philosophy that bridges the gap between realism, in which an objective external reality is posited, and constructivism, in which reality is crafted entirely by epistemic agents. She constructs knowledge as both relational and distributed among the epistemic agents interacting within a synthetic socio-technical system; it emerges from the collaboration of all agents capable of processing information. These agents can influence the environment, introducing new dynamics between it and other agents (Russo, 2022). Thus, knowledge is always contingent upon the specific configuration of relations with a synthetic socio-technical system, both between the world and the epistemic agents but also among those agents themselves. Russo further elaborates on this notion of relational knowledge, stating:

> Knowledge is relational also at the level of the concepts or of the semantic artefacts that compose it. I speak of semantic artefacts to emphasise that we make the concepts that we use to make sense of the world around us. I explored the idea that human epistemic agents are makers not just because we make artefacts, but also because we make semantic artefacts or concepts. To say that knowledge is relational at the levels of concepts means that these are not islands but are always connected to other concepts. I take this to be an irreducible relational aspect of knowledge" (Russo, 2022, p. 163).

In our contribution, we take the pivotal insight that knowledge, including concepts, is a kind of artefact that stems from the epistemic collaboration between poietic agents, both human and non-human. Due to their ability to process information, Russo argues that artificial agents qualify as genuine epistemic agents that display poietic capabilities and thus are capable of shaping the dynamics of the socio-technical system in which they exist. This is the sense in which humans interact *with* generative AI: humans and AI systems co-produce semantic artefacts. Consequently, artificial agents alter the overall configuration of relations and contribute to this collaborative production of knowledge. This notion of the poietic agent allows our theoretical analysis to delve deeper into the dynamics of synthetic socio-technical systems and to elucidate their functionality and implications in a way that ANT cannot. To anticipate our argument, it is only through the concept of poietic agent that we can spell out the role of artificial agents as *meaning makers* (hence the relevance of semantic artefacts and of knowledge) in synthetic socio-technical systems. However, it is important to note that possessing poietic abilities, i.e. being meaning-makers, is not something that agents either have or do not have, but is a capability that is exhibited in degrees. Arguably, analogue technologies participate in the process of co-production of knowledge, but not in the form of meaning making. Furthermore, while digital technologies such as a Microsoft Word displays higher capability of processing information than old-fashioned typing machines, they are less capable than newer versions of this software that integrate generative AI tools for writing.

Having introduced the concept of poietic agent, it is useful to introduce some qualifications about artificial agents, AI being the relevant divide. The distinction between 'old-fashioned' artefacts and artificial agents lies in the latter's ability to autonomously generate new semantic artefacts. Previously, technologies in socio-technical systems primarily played the role of conveying semantic artifacts that were produced by human agents (discourses, narratives, etc.). This conveying activity was by no means neutral, as ANT explains. Nevertheless, the level of autonomy in altering the information, and thus social relations, was qualitatively limited. At best, an artifact was capable of altering the message in its ability to reach other agents, its 'magnitude', and only minimally altering its content, and only because of entropy. This could occur in the ways described by Latour, essentially an imprecise translation of the original message, with little or no ability to calibrate this change based on a specific semantic intention. This, of course, does not imply that 'old-fashioned' artifacts did not have a considerable influence on social processes. Altering the 'reach' of certain types of semantic artifacts, as Facebook did by amplifying some communication modes at the expense of others, produces significant social effects. By contrast, the level of autonomy showed by AI systems, and specifically generative programs, is unprecedented and is rapidly increasing.

The proposed concept of poietic agent does not prioritise human agents in the production of knowledge and so constitutes an appropriate theoretical approach for our current socio-technical situation in which artificial agents are part and parcel of our daily socio-technical systems, that are thus becoming *synthetic*. As discussed, previous information communication technologies were, for the most part, characterized as objects that humans interacted *through* and so behaved as *meaning-mediators* that only conveyed those semantic artifacts produced by human agents (e.g., discourses,

narratives). With the arrival of generative AI systems and most famously LLMs, artificial agents in our social environment are now in a position to autonomously generate new semantic artifacts with human agents only partially contributing to their creation, such as in the selection of training data, and thus qualify as meaning *makers* (poiêsis of semantic artefacts). While it is incorrect to argue that generative AI systems are capable of creating semantic artifacts *ex nihilo*, the degree of autonomy that these programs have in the semantic production process has dramatically increased. Not only do AI systems construct new semantic artifacts that do not require human intervention to be valid, but they can further adjust their semiotic outputs based on the social environment and can, to a certain extent, anticipate how these semantic artifacts will be integrated into and thus affect the social environment. It is important to emphasize that these programs do not understand the meaning of the semantic artifacts they produce as they operate according to formal semantic relations and probabilistic principles. Nevertheless, the semantic artifacts they produce, either independently or in collaboration with human agents, become integral to what semiotician Juri Lotman terms the *semiosphere* (Lotman, 2022). This semiosphere refers to the overall reservoir of meaningful symbols, signs and narratives that inform our cultural and social comprehension and is understood as something that continually evolves with the emergence of new semantic artifacts. Even though these artificial agents do not understand the meaning of the semantic artifacts they produce, they now inhabit our world as *meaning-makers* that contribute to the semiosphere and are thus changing it in ways previously inconceivable.

Now let's make another step in our argument: it is not simply that the arrival of generative AI systems will vastly increase the number of poietic agents producing semantic artifacts but that the semantic artifacts they produce are fundamentally different from those produced by human agents. Here, we do not delve into the intricate workings of social systems, their socio-symbolic structures, how they achieve equilibrium, or why they evolve. The interested reader is referred to Bisconti (2024) for a discussion of such issues. Our primary concern is to discern how artificial agents might introduce a *qualitative* shift in the organization of social systems, leading to the emergence of synthetic systems. We suggest that the transformative potential of generative AI lies in the 'imperfections' of Deep Neural Networks (DNNs), the foundational technology behind this form of AI. As evidenced by research in computer science and AI, DNNs often lack the precision and stability desired (del Campo & Leach, 2022; Ji et al., 2023). If generative AI were capable of producing artifacts as flawlessly as humans, one could argue that synthetic socio-technical systems would simply be traditional social systems with a vastly increased number of agents creating semantic artifacts. This would already signify a major shift in how social systems are organized. However, the full impact of AI extends beyond this. The imperfections in AI processing can lead to outputs that are not only unexpected but also significantly different from the training data they aim to replicate. This introduces a new dimension of variability and unpredictability in AI-generated content. Furthermore, if these programs are trained using synthetic datasets (e.g., AI-generated text or images) as well as or instead of authentic data, their outputs will progressively deviate further and further away from human outputs. As such, those semantic artifacts generated

by AI systems can differ drastically from those human artifacts they are intended to mimic.

With this argument we aim to emphasize it is paramount to enquire to what extent synthetic socio-technical systems should be considered different and unique from previous socio-technical systems. The reason is that semantic artifacts generated by AI differ from human artifacts and, at the same time, these programs are becoming more deeply integrated into society. This uniqueness entails a reflection on which kind of policy responses might be able to manage this new form of socio-technical systems. In the next section, we will discuss some policy implications of our paper.

5 Conclusions

The recent arrival and widespread integration of generative AI systems across society is instigating a profound change in our social environments as we now have technologies that we interact *with* in a similar way we interact with other humans. Due to the pervasive use of generative AI in social contexts, this paper argues that our traditional socio-technical systems have thus become *synthetic* in nature; we use 'synthetic' to express the idea that the social interactions within such systems cannot be considered as something pertaining only to humans, nor are they mediated in the same way as 'old fashioned' artifacts. Generative AI systems are poietic agents able to generate semantic artifacts in such a way that they are able to influence social systems in the same (semantic) way humans do. However, it is particularly difficult to analyse the social impact of these new synthetic socio-technical systems because current frameworks do not fully account for AI as a social actor with proper poietic characteristics.

Thus, this paper introduced a system-level approach that builds on ANT and on the philosophy of information in that it views human and artificial agents as social actors equally capable, from the qualitative point of view, of influencing the social environment, since they are now able to create new semantic artifacts. Obviously, the quantitative degree to which they are able to shape social systems is, as for today, still not comparable. At the same time, AI systems do generate some semantic content that becomes thus part of the 'semiosphere'. For this reason, we think that the concept of poietic agent allows us to begin moving toward a more accurate evaluation of human-AI social interactions that uses appropriate criteria.

This paper is primarily theoretical and seeks to establish a conceptual framework for analysing the social implications of emerging AI systems. In this concluding section, we aim to derive the relevance of this framework for policy, while it is beyond the scope of this work to develop extensive and detailed policy recommendations.

Firstly, the concept of human-centered AI, as promoted in policy discussions, will take on a profoundly different meaning if we recognize that social systems are almost equally shaped by both humans and non-humans. Typically, human-centered AI implies that these systems are tools—potentially highly intelligent ones—that support human activities. This perspective suggests that the proliferation of AI systems is not fundamentally different from that of other technologies, implying that social systems will remain largely unchanged in their operational dynamics despite the widespread adoption of a new technological tool. Contrarily, we argue that generative AI

has distinct capabilities that differentiate it from traditional technologies. Thus, AI systems cannot be merely designed as human-centered tools since they modify the social environment by creating semantic artifacts, which play a crucial role in social interaction. This redefines the interaction dynamics within social systems, challenging the traditional notion of technology merely supporting human activity.

The second policy implication of adopting our framework is the valuable opportunity to apply Actor-Network Theory (ANT) to analyze how AI systems not only create new semantic artifacts but also modify existing ones. Additionally, through an ANT analysis, one can identify multiple entry points to design policies, at different joints of the network. This approach can deepen our understanding of the other actors involved in shaping these artifacts. Additionally, the production and dissemination of these 'artificial' semantic artifacts merits sociological investigation to explore their integration into and impact on the existing 'semiosphere.' Generative AI has the capability to seamlessly contribute to our collective knowledge, much like human-created artifacts.

However, the techniques employed by generative AI to generate these new semantic artifacts often display unique idiosyncrasies that differ from traditional human methods of producing semantic content. Investigating these peculiarities could offer researchers valuable insights into potential evolutionary paths for social systems. This is particularly relevant as an increasing proportion of semantic artifacts—whether in text, image, audio, or other media formats—are being produced by AI systems. Understanding these dynamics is crucial for shaping policies that recognize the profound role AI plays in the ongoing evolution of social systems.

This consideration leads us to the third area of policy inquiry, which is interconnected with the others: what does trustworthiness mean in the context we have described? Traditionally, trustworthiness is an attribute associated with human beings, conceptualized through benevolence, competence, and integrity (Burke et al., 2007). This framework continues to be prevalent in discussions about AI trustworthiness. However, recent studies, such as those by Aquilino et al. (2024), are beginning to challenge this direct transfer of human trust paradigms to AI systems. How should we reconceptualize social trust in environments where social systems are synthetic? The introduction of non-human agents presents novel challenges in how trust is constructed among actors in social systems.

Clearly, the policy implications of synthetic socio-technical systems warrant a more comprehensive discussion, which is beyond the scope of this paper. Given the current absence of a robust theoretical framework for fully understanding these issues, this paper has concentrated on conceptual discussions. Future work will more thoroughly explore the policy implications derived from our framework.

Author contribution All authors contributed equally to this manuscript.

Data Availability Given the purely theoretical nature of this paper, Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Declarations

Ethical Approval The research did not involve any experiment requiring ethical approval.

Consent to Participate The research did not involve any participant.

Consent to Publish All authors gave consent to publication of this manuscript.

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

Aquilino, L., Bisconti, P., & Marchetti, A. (2024). Trust in AI: Transparency, and uncertainty reduction. Development of a new theoretical framework. In CEUR workshop proceedings (pp. 19–26). CEUR-WS.org.

Bisconti, P. (2024). Hybrid Societies: Living with Social Robots. Routledge. https://books.google.at/books?id=FDQw0AEACAAJ.

Bisconti, P., & Carnevale, A. (2022). Alienation and recognition: The Δ phenomenology of the Human-Social Robot Interaction. *Techne: Research in Philosophy and Technology*, *1*.

Bommasani, R., et al. (2022). *On the opportunities and risks of foundation models*. Centre for Research on Foundation Models.

Burke, C. S., Sims, D. E., Lazzara, E. H., & Salas, E. (2007). Trust in leadership: A multi-level review and integration. *The Leadership Quarterly*, *18*(6), 606–632.

Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, *98*, 147.

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., & Gehrmann, S. (2022). Palm: Scaling language modeling with pathways. *ArXiv Preprint ArXiv*:220402311.

Chun, W. (2021). *Discriminating data: Correlation, neighbourhoods, and the new politics of recognition*. MIT Press.

Coeckelbergh, M. (2011). You, robot: On the linguistic construction of artificial others. *AI and Society*, *26*(1), 61–69. https://doi.org/10.1007/s00146-010-0289-z

Coeckelbergh, M. (2016). In G. David, F. J, & M. D (Eds.), *Alterity ex Machina: The Encounter with Technology as an epistemological-ethical drama* (pp. 181–196). Rowman & Littlefield.

Coeckelbergh, M. (2023). Democracy, epistemic agency, and AI: Political epistemology in times of artificial intelligence. *AI and Ethics*, *3*(4), 1341–1350.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.

del Campo, M., & Leach, N. (2022). *Machine hallucinations: Architecture and Artificial Intelligence.* Wiley.

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, *23*(7), 2072–2098.

Duffy, B. R., Rooney, C., O'Hare, G. M. P., & O'Donoghue, R. (1999). What is a social robot? 10th Irish Conference on Artificial Intelligence \& Cognitive Science, University College Cork, Ireland, 1–3 September, 1999.

Ezenkwu, C. P., & Starkey, A. (2019). Machine autonomy: Definition, approaches, challenges and research gaps. In Intelligent Computing: Proceedings of the 2019 Computing Conference, Volume 1 (pp. 335–358). Springer International Publishing.

Floridi, L. (2011). A defence of constructionism: Philosophy as conceptual engineering. *Metaphilosophy*, *42*(3), 282–304.

Floridi, L. (2013). *The ethics of information*. Oxford University Press.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.

Floridi, L., & Sanders, J. W. (2005). Internet ethics: The constructionist values of homo poieticus.

Gunkel, D. (2017). *The changing Face of Alterity*. Rowman & Littlefield.

Hempelmann, C. F., Raskin, V., & Triezenberg, K. E. (2006). Computer, tell me a joke… but please make it funny: Computational humor with ontological semantics. FLAIRS 2006 - Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference, 2006(1994), 746–751.

Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth*. Indiana University Press.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1–38.

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*(MAR), 1–16. https://doi.org/10.3389/fpsyg.2015.00390

Latour, B. (1996). On actor-network theory: A few clarifications. *Soziale Welt*, 369–381.

Latour, B. (2002). Una sociologia senza oggetto? Note sull'interoggettività. *LANDOWSKI E.; MARRONE G. La Società Degli Oggetti. Problemi Di Interoggettività. Roma: Meltemi*.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.

Latour, B., & Law, J. (1992). (MIT Press, Cambridge, MA) Pp, 254–258.

Lotman, J. M. (2022). *La Semiosfera*. La Nave di Teseo Editore spa.

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, *7*, 33–35.

Naneva, S., Sarda Gou, M., Webb, T. L., & Prescott, T. J. (2020). A systematic review of attitudes, anxiety, Acceptance, and trust towards Social Robots. *International Journal of Social Robotics*, *12*(6), 1179–1201. https://doi.org/10.1007/s12369-020-00659-4

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of human behavior in human - Robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE Transactions on Robotics*, *24*(2), 442–451. https://doi.org/10.1109/TRO.2007.914004

Rini, R. (2020). Deepfakes and the Epistemic Backstop. *Philosophers' Imprint*, *20*(24), 1–16.

Ropohl, G. (1999). Philosophy of socio-technical systems. *Techne: Research in Philosophy and Technology*, *4*(3), 186–194. https://doi.org/10.5840/techne19994311

Russo, F. (2022). *Techno-scientific practices: An informational approach*. Rowman & Littlefield.

Seibt, J. (2017). Towards an ontology of simulated social interaction: Varieties of the as if for robots and humans. *Sociality and normativity for robots* (pp. 11–39). Springer.

Simondon, G. (2020). *Individuation in light of notions of form and information*. University of Minnesota Press.

Stock, O., & Strapparava, C. (2003). Getting serious about the development of computational humor. IJCAI International Joint Conference on Artificial Intelligence, 59–64.

Trist, E. L. (1981). *The evolution of socio-technical systems* (Vol. 2). Ontario Quality of Working Life Centre Toronto.

Verbeek, P. P. (2016). *Toward a theory of Technological Mediation A Program for Postphenomenological Research* (pp. 189–204). Technoscience and Postphenomenology. The Manhattan Papers.

Walters, M. L., Syrdal, D. S., Dautenhahn, K., te Boekhorst, R., & Koay, K. L. (2008). Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots*, *24*(2), 159–178. https://doi.org/10.1007/s10514-007-9058-3

Wang, X., Shen, J., & Chen, Q. (2022). How PARO can help older people in elderly care facilities: A systematic review of RCT. *International Journal of Nursing Knowledge*, *33*(1), 29–39.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11).