AI & SOCIETY
https://doi.org/10.1007/s00146-024-02037-4

OPEN FORUM

# Entangled AI: artificial intelligence that serves the future

Alexandra Köves[1] · Katalin Feher[2] · Lilla Vicsek[3] · Máté Fischer[4]

Received: 14 December 2023 / Accepted: 23 July 2024


**Abstract**
While debate is heating up regarding the development of AI and its perceived impacts on human society, policymaking is struggling to catch up with the demand to exercise some regulatory control over its rapid advancement. This paper aims to introduce the concept of entangled AI that emerged from participatory backcasting research with an AI expert panel. The concept of entanglement has been adapted from quantum physics to effectively capture the envisioned form of artificial intelligence in which a strong interconnectedness between AI, humans, society, and nature is reflected. Entanglement assumes that AI should serve nature, social well-being, justice, and the resilience of this intertwined network simultaneously and promote a dynamic balance among these factors. This approach allows us to understand the pervasive role of this technology and the scope of human agency in its development. The study shows how such concepts seem to transcend the dominant discourses related to expectations, technological determinism, and humanism. An additional aim of this paper is to demonstrate how backcasting can contribute to generating useful understandings of the future of AI and fruitful insights for policymaking.

**Keywords** Artificial intelligence · Backcasting · Entangled AI · Expectations · Interconnectedness · AI ethics

## 1 Introduction

A highly referred open letter was signed in April 2023 by prominent businesspeople and scientists calling for a pause in the development of artificial intelligence to keep the technology under control. Likewise, a systematic literature review (Feher and Katona 2021) uncovered the issues of technological trustworthiness with broad societal impacts, but many are also concerned with the ability to regulate it adequately, especially with transparency (James et al 2023; Andrada et al. 2023; Larsson and Heintz 2020). Hence, the demand for understanding the impacts and reducing the potential risks of AI is well founded, but an insight into future concepts supports an even broader perspective.

The aim of this paper is to demonstrate how backcasting can contribute to generating valuable understandings of the future of AI and provide fruitful insights for policymaking. We borrow the concept of entanglement from quantum physics to effectively capture the envisioned form of artificial intelligence in which a strong interconnectedness between AI, humans, society, and nature is reflected. Entanglement assumes that AI should serve nature, social well-being, justice, and the resilience of this intertwined network simultaneously and promote a dynamic balance among these factors. This approach allows us to understand the pervasive role of this technology and the scope of human agency in its development. Hence, it facilitates the convergence of social science and technological development perspectives.

Expectations connected to technologies have been argued to be of key importance in modern societies (Borup et al

✉ Alexandra Köves
alexandra.koves@uni-corvinus.hu

Katalin Feher
katalin.feher@uni-nke.hu

Lilla Vicsek
lilla.vicsek@uni-corvinus.hu

Máté Fischer
mate.fischer@stud.uni-corvinus.hu

1 Department of Decision Sciences, Institute for Operations and Decision Sciences and Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Fővám tér 8. E143, 1093 Budapest, Hungary

2 Department of Social Communication, Ludovika University of Public Service, Ludovika tér 2, 1083 Budapest, Hungary

3 Department of Sociology, Institute for Social and Political Sciences, Corvinus University of Budapest, Fővám tér 8. E143, 1093 Budapest, Hungary

4 Department of Decision Sciences, Institute for Operations and Decision Sciences, Corvinus University of Budapest, Fővám tér 8. E143, 1093 Budapest, Hungary

Published online: 09 August 2024

2006; Brown and Michael 2003) with a definition of "statements about future conditions or developments that imply assumptions about how likely these are supposed to be and that travel in a community or public space" (Konrad et al. 2016: 65). Thus, expectations include the evaluation of likelihood or plausibility with a performative role to legitimize and coordinate the actions of a range of actors and attract investment (Van Lente 2012). However, expectations often fail, not only because of encouraged investments instead of accuracy but also because many social factors are left out when they are formulated (Geels and Smit 2000).

Individual approaches to the future of AI have been grouped by Makridakis (2017) into four types. Optimists expect AI to enable a utopian future of abundance for humans, where people do not need to work but instead do activities they prefer, and technology, in general, makes life better. Pessimists fear a dystopian future with AI, where humans face poverty, inferiority, or even extinction. Pragmatists are of the view that AI does not pose a danger as long as adequate regulation is in place or controls are built in to make machines stop working if threatening situations arise. Doubters assert that AI will not be able to do much compared to humans anyway and thus does not pose a danger.

This thinking about AI futures via expectations has often been characterized by technological determinism (Mackenzie and Wajcman 1999). Technological determinism implies that technologies change either due to neutral scientific advances or due to an inherent logic of their own, and only then do they impact society. Hence, technology is recognized as neutral, apolitical, and independent of society such as an inward-directed mechanical force derived from a distinct entity (Beckert 2016). It disregards impacts on power, culture, or social factors and the role of human agency in shaping technology. This approach is heavily criticized in science and technology studies. It has also been argued that this kind of focus invalidates democratic deliberation, leaving politics and society to the role of just adapting, rather than shaping technological change [Dotson (2015)].

Besides these dominant technologically deterministic narratives which view AI as neutral, alternative narratives are gravely needed to reframe how the future is envisioned. To some degree, this work has already started. In recent years, growing attention has been paid to why AI is not neutral—for example, with regard to the inbuilt biases of different algorithms (Zou and Schiebinger 2018). AI ethics has become a flourishing field, even though it can be argued that ostensibly paying attention to ethics may be used by companies as a way of avoiding further regulation of the field and pacifying the public (Kerr et al. 2020). AI initiatives have also been created as part of the growing field of technological development entitled 'Technology for Good' (Powell et al 2022).

In this article, we illustrate that it is not just the expectations associated with assessments of probabilities but also the preferred visions of the future that can be performative: they have the potential to play an important role in modern societies (Vicsek 2021). A case in point is the sustainability discourse that results in the formulation of different visions about preferred sustainable futures, which have influenced many forms of technological development. We argue that taking into consideration preferred visions of artificial intelligence is just as crucial for understanding the intertwined nature of technology and society and establishing adequate societal and policy responses to its development.

Backcasting (Robinson 2003) is a method of creating preferred visions and proposing policies for how we might arrive at them. We posit that backcasting and systems mapping (Barbrook-Johnson and Penn 2022) can be useful tools for advancing thinking about 'AI for good' as they can benefit policy and strategy development. Rather than relying on existing projections of the future as the basis for policy development, it can be useful if decision-makers consider alternative forms of development. Backcasting can help generate these alternative narratives, as participants are asked to imagine an ideal future and think of the activities that could lead to it. If the results are considered by decision-makers, they can help guide outcomes in a preferred direction. While other methods of foresight often assume that actors simply adapt to trends and events, including technological change, backcasting assumes that stakeholders can deliberately create a desired future (Robinson 2003).

We conducted a backcasting research involving a panel of AI experts panel in Hungary in Summer 2022. Participants came from various backgrounds, including the business sphere, academic sphere, and civil society. The study thus makes a unique contribution to the social thinking about AI futures, as envisioning the preferred future of artificial intelligence by applying backcasting methodology is a neglected topic within the literature on AI. From a search of the top-ranked journals, we could identify no backcasting projects that looked at the normative future of AI in general, and just a few cases when AI was investigated in a narrow context, being mentioned as just one of the important factors to be examined (Lee 2023).

The research being of an abductive nature (Timmermans and Tavory 2012), during the analysis of the backcasting results, the concept of entangled AI emerged. Participants envisioned AI as strongly interconnected with humans, society, and nature and as a technology that humans have the capacity to shape. Therefore, as unexpected research insights, we also introduce the conceptual framework of an interconnected 'entangled AI'. The participants themselves did not use this phrase, but for reasons of appropriate and concise interpretation, we introduce this concept as it reflects well the common understanding of the participants of what AI should be in the future.

Our paper consists of six sections. First, we address the issue of interconnectedness in the literature. In the methodology section, we discuss backcasting in general and our research in more detail. In the subsequent section, we describe the results and depict the normative systems view of entangled AI. The discussion connects our findings to the debates on interconnectedness and humanism/posthumanism and discusses the main messages for policymaking. Finally, a short overview of the implications is offered in the conclusions.

## 2 Interconnectedness

Entanglement in quantum physics is a "subtle, beguiling quantum phenomenon that embodies the radical interdependence of things" (Rovelli 2022). In detail, entanglement involves superposition, where quantum systems exist in multiple states at once, entanglement correlating systems, and non-locality, linking distant quantum systems (Gordon 2023). In various other fields, the concept of interconnectedness appears more extensively and is closely connected to the framing of entanglement, i.e., the deep complexity of a given phenomenon. For example, intertwined environmental and organizational factors (Van Noordt and Misuraca 2022), self-optimizing production systems (Permin et al 2016), the interdependencies of critical infrastructures (Haimes 2002), and reproduced social inequalities (Diakopoulos et al. 2017) are analyzed this way in a global context and with an algorithmic culture. In this framing, AI is a form of supernatural and liberated agency (Floridi 2023) and related ethical issues can be studied through socio-technical entanglement (Herzog and Diebel-Fischer 2022; Draude et al 2020). Barad (2007) argues that the concept of entanglement, which involves the lack of an independent, self-contained existence, offers an ontological, epistemological, and ethical approach to understanding the complex interconnections between humans, society, and nature on a global scale.

The entanglement metaphor parallels the assemblage concept in science and technology studies, which also underscores the multifaceted interconnections within socio-technical systems (Kubes and Reinhardt 2022). Some social scientists even use the concepts of entanglements and assemblages interchangeably (Kautz and Blegind 2012). Assemblage theory, particularly as conceptualized by Deleuze and Guattari (1987), views reality as composed of heterogeneous elements that dynamically interact and connect to form complex wholes. It emphasizes the idea of interconnectedness by highlighting how these diverse components influence and co-constitute each other within an assemblage, creating unique, fluid, and non-hierarchical relationships However, entanglement in some understandings implies a deeper, intrinsic interconnectedness, where the alteration of one element instantaneously and fundamentally affects the others (Rovelli 2022). This nuanced understanding is critical for AI, as it encapsulates the simultaneous, multi-directional influences between AI, society, and nature, highlighting the profound consequences of technological interventions. Furthermore, the quantum notion of entanglement emphasizes the idea of non-locality, where distant entities remain interconnected in ways that defy classical understandings of space and time. This aspect resonates with the global and instantaneous impact of AI, transcending geographical and temporal boundaries (Floridi 2023).

Kubes and Reinhardt (2022) propose an approach to constructing a relational ontology of multi-species assemblages, from robots to humans. In such an interpretation, humans are only nodes in this interconnectedness. Even if we presume that conscious humans are at the center of AI development (Keeling and Lehman 2018), from such a systemic perspective, they are also ultimately influenced by a complex web of societal, technological, and ecological factors. On a philosophical level, arising issues of interconnectedness and human agency can also be assessed in relation to the discourse on humanism and posthumanism. While humanism treats humans as worthy, superior, and capable agents of generating change, posthumanism emphasizes interconnectedness with physical, chemical, technological, and biological environments (Herbrechter et al.2022). According to posthumanism, humans do participate in bringing about change, but there are different views concerning whether "a conscious human subject can actively create change" (Keeling & Lehman 2018). With respect to interconnectedness, posthumanist thought shares some characteristics with certain types of indigenous thought stemming from a much earlier era (Bignall 2016). Illustrating with an example, Maori philosophy supposes that "all matter is profoundly interconnected" (Irwin and White 2019) and that "human beings are one species embedded in a network of familial relations with ecological place. There is no separation or alienation of people from the land". This kind of view not only argues for interrelatedness but dethrones humans as the center of the universe, which ties in with posthumanist philosophical approaches (Kriman 2019). Posthumanism similarly adopts this stance, perceiving humans not as the pinnacle of value but as part of a broader, interconnected existence. Complementarily, recent advancements in AI research echo this sentiment, forecasting a human–machine ecosystem that draws inspiration from natural systems (Feher et al. 2024). While this naturalistic ontology, Descola (2014) represents merely one avenue for framing the relationship between humans and their environment, the exploration of these entanglements plays a pivotal role in deciphering the complexities of

**Table 1** Background of participants

| Background | Sector |
|---|---|
| AI developer at a multinational corporation in the energy industry (the only female participant) | Business |
| Data scientist working at a multinational corporation in the electronics industry | Business |
| AI developer at a multinational corporation in the automobile industry | Business |
| AI expert, member of the AI Coalition, working at an SME specializing in the practical application of Artificial Intelligence (AI) for enterprises and other organizations | Business |
| Lawyer specialized in AI legal consultancy | Business and Academia |
| IT specialist running educative community projects especially for teenagers and young adults | Business and civil society |
| IT specialist, head of a non-profit organization promoting open-source knowledge | Civil society |
| IT scholar, head of the Department for Artificial Intelligence at a leading Hungarian university | Academia and Business |
| Associate professor at an Institute for Business Informatics at a leading Hungarian university working in the field of disruptive technologies | Academia |
| Ecological economics professor at a leading Hungarian university doing research on technology's societal and environmental impact | Academia |
| AI specialist focusing on AI applications in the healthcare system working at a leading medical university operating a large number of clinics in Hungary | Academia |
| AI developer at a research hub at a leading Hungarian technology university | Academia |
| Economic sociologist at a leading Hungarian university specialized in social innovation | Academia |
| PhD student doing research in the field of AI | Academia |
| PhD student doing research in the field of AI | Academia |
| Journalist specialized in covering AI issues | Media |

socio-technological landscapes, thus enriching the discourse within the social sciences.

## 3 Methodology

Nowadays, human societies are facing many complex and systemic problems for which it is particularly difficult to find answers using the existing paradigms. The normative framework of backcasting attempts to create an ideal future scenario about the phenomenon being investigated and then elaborates on the activities required to get there (Quist and Vergragt 2006). Repositioning the understanding of our realities further into the future creates room in individuals' minds for creative thinking and contemplation, undistorted by the current path dependencies and lock-in effects (Köves and Király 2021).

While forecasting is intended to provide projections of the most probable future circumstances considering current conditions as a starting point, backcasting turns this logic upside down (Robinson 2003). Backcasting is particularly effective in circumstances when the future is uncertain, the heterogeneous systems involved are complex, and the outputs that can be predicted based on the current trends are undesirable and suitable for developing alternative future visions in broad subject areas for longer periods of time that go beyond mainstream frameworks. The role of artificial intelligence in the service of the future is exactly such a highly complex problem to which the chosen methodological approach fits well.

Due to the enormity of the topic and its broad social implications, our research was based on participatory backcasting (Bergold and Thomas 2012). Participatory backcasting implies that future scenarios are not created by researchers but by the participatory process itself.

We opted for a panel the members of which were somewhat connected to artificial intelligence in particular or technology in a broader sense through their work. The diverse backgrounds of the participants covered many fields, from the business sphere through civil enterprises to the world of academia. Table 1 provides the short description of the participants' background. While we managed to secure diversity in terms of sector, despite all our efforts, the project did not meet race and gender diversity criteria as participants were all male except for one and White. This situation, unfortunately, reflects the diversity of people involved in the development of AI in Hungary.

As Table 2 shows, the backcasting process associated with the research—for which a 2 × 1-day workshop in the summer of 2022 created the backbone—may be divided into four main phases.

On the first day of the workshop, the participants were asked to come up with ideal visions of the future using short headlines based on previously defined themes using the World Café approach (Brown 2010). Participants were invited to think creatively and associatively in changing groups. Before the second workshop, participants reviewed

**Table 2** The process of our backcasting research—own construction

| Backcasting stage (based on Quist et al. 2011) | Research methodology | Outputs | Timeframe (net time) |
|---|---|---|---|
| 1. Strategic problem orientation | Brainstorming with workshop techniques (Putman and Paulus 2009) | Future vision fragments | 1 h |
| 2. Future vision development | World Café (Brown 2010) | The future vision itself | 5 h |
| 3. Backcasting analysis | Futures Wheel (Glenn 2009)—slightly modified | Identified backcasting steps (tools and recommendations for intervention) | 4 h |
| 4. Future alternatives | Systems mapping with the casual loop diagram technique (Barbrook-Johnson and Penn 2022) | Key variables and relationships between them | 10 h after the workshop |

the vision synthesized by the researchers, and the final form of the text was born through iterations. During the second workshop, participants in small groups defined goals related to four key areas for 2060. The intervention steps that would support the achievement of the final goals were first defined for 2050, 2040, and 2025.

To reveal the connections that emerged behind the elements of the visions, the goals, and the intervention steps, the researchers decided to visualize the logic of the argumentation of participants in a systems' map (Barbrook-Johnson and Penn 2022). While systematizing the results of the backcasting, the researchers defined the variables of the systems map by analyzing the text of the future vision as well as the goals and intervention steps suggested by the participants. This was followed by the mapping process, during which the relationships between the categorized variables and the reinforcing (positive) and balancing (negative) feedback loops were visualized (chart 1) using the argumentation of participants as the source.

More details of the process, the full text of the vision, as well as a table describing all the variables of the system map can be found in the research report (Köves et al. 2023).

## 4 The normative systems view of entangled AI

Participants in the backcasting research envisioned artificial intelligence to be part of a complex system of intertwined natural, human, and technological environments, where all constituent elements interact with each other, influence each other, and rely on each other. In the vision, this interdependence of AI and the surrounding world is described in terms of its relationship to humans, the conceptual world, and the physical world, including social and ecological realities.

The vision states that *"AI does not exist independently, but in close interaction with humanity, as an extension of a kind of global collective consciousness, reflecting people's ethical value choices"*. The discourse involved an AI that almost mirrors the consciousness of human beings; AI is what we are. If we use AI for profit-seeking, AI will do its best to regenerate the circumstances and solutions that lead to profit for those who seek this. If we perceive that the world is inherently unjust and unequal, AI will also treat it that way. If human consciousness can transcend the dominant narratives of our times, AI will follow suit. The perception was that AI is capable of interweaving our lives in many different ways, but its use will always be constrained by the value choices that we humans make.

The other perspective of entanglement is the interdependent relationships of the ecological, social, and economic spheres and the AI that is interlaced throughout them. Participants adopted a strong systems view (Capra and Luisi 2016) of the future, as is well reflected even in the first paragraph of their vision: *"We regard artificial intelligence as part of our environment, like the ecological environment, and strive to live in a harmonious relationship with both…From this interdependence, it follows that the environmental burden cannot exceed the carrying capacity of our planet either in the case of human activity or artificial intelligence."*

The systems map (Fig. 1) is a good visualization tool for replicating this kind of interdependence, even if simplified. However, although the researchers drew the systems map, not the participants, all efforts were made only to illustrate the relationships that were discussed during the backcasting workshop. Nonetheless, the map contains a sufficient number of variables and relationships to provide us with a good overview of the qualitative participatory research.

### 4.1 The aim of entangled AI

The most central variable in the map is individual well-being, i.e., the subjective assessment of individual quality of life. This is a measure of how an individual feels about life beyond the dimensions of material well-being. Ten variables directly influence *individual well-being*, but there was no discussion about what influences the latter dimension for the simple reason that participants seem to have treated individual well-being as the ultimate aim of the whole system. The ecological environment would

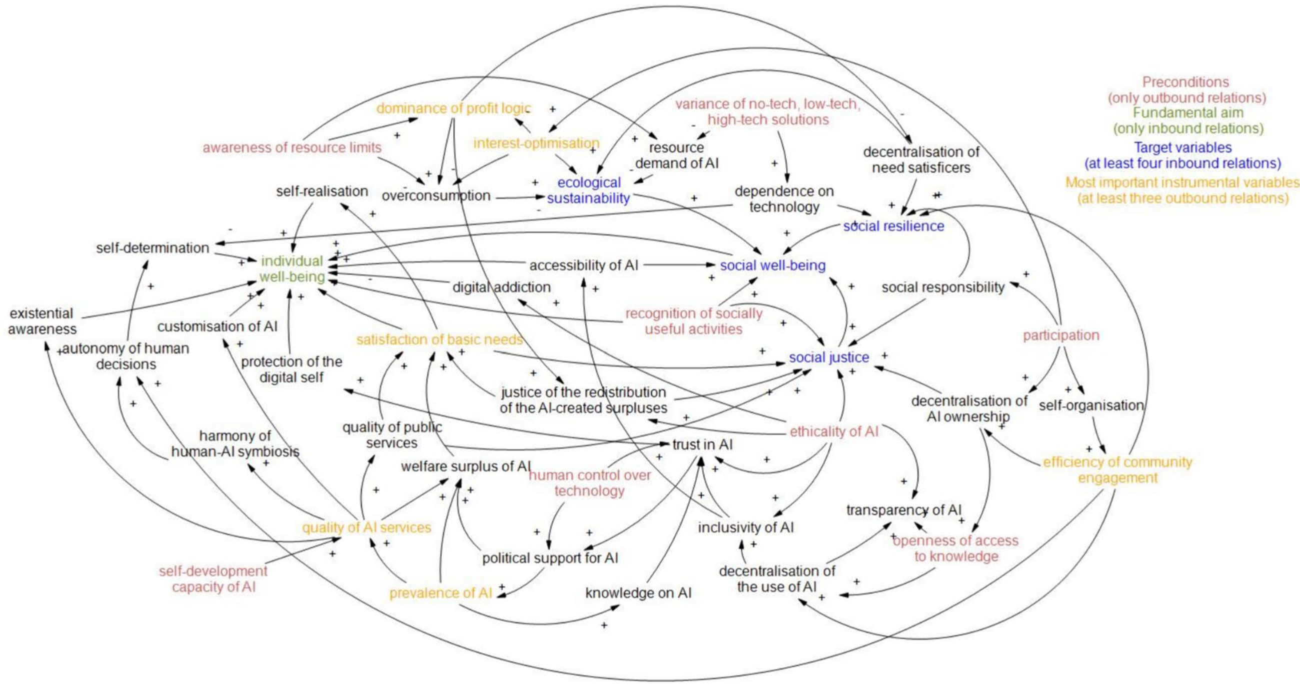


**Fig. 1** Systems map of entangled AI—authors' illustration based on the backcasting research

not need AI if humans did not exist, and—in an ideal scenario—societies would not need AI if it were not engineered to promote the well-being of every member of that society.

Four variables also have at least four inbound relationships in the map: ecological sustainability, social justice, social well-being, and social resilience. These can be considered the areas identified by the participants where AI may contribute the most to the overall function of the entangled system. For example, without maintaining ecological conditions, any technological advancement would be useless as the very sustenance of living conditions would be jeopardized. However, our ecological environment is such a complex system, and human activities directly and indirectly impact it in such entangled ways that no human can oversee these diverse interconnections and monitor changes in the system. On the other hand, AI can support decisions that involve human impacts on nature and monitor changes in nature that impact human societies and keep track of balances and trade-offs on individual and collective levels.

The other three target variables in the system concern the vision of the participants regarding social conditions, where justice, collective well-being, and resilience were in the focus of the discussions. According to participants, AI is capable of supporting all these social phenomena (well-being, justice, resilience) as long as people themselves understand the importance and relevance of these concepts.

## 4.2 Preconditions for entangled AI

This is where we come to another set of interesting variables in the system of entangled AI: the preconditions. Variables that only have outbound relations and seemingly remain disembedded in the systems map cover factors that participants considered prerequisites for the whole system to operate. They are like credos of participants, whereas questioning the importance of these issues would likely result in alternative visions.

Participants agreed that humanity has to respect Earth's regenerative capacities. AI can promote respect for limited resources by facilitating both the acceptance and our knowledge of our physical limits. In the case of entangled AI, technology will only support harmony with nature as far as humans themselves act consciously in this direction. As an example AI providing humans with information on their carbon footprint and potential reduction steps only makes sense if it matters and we follow the AI's suggestions.

Participants also presumed that in the future solutions to given problems would range from no technological intervention, through just basic equipment, to complex technologies. An essential condition for an entangled AI is that humans continue to oversee and influence the development trajectories and guiding principles of AI, e.g., through expert communities. This control happens, even though participants presumed that AI would be fully capable of developing itself. Nonetheless, both human influence over the trajectories and AI's self-development must reflect explicitly deliberated ethical choices that play a significant and ongoing role in accepting or declining technological solutions.

This also means that entangled AI can only support individual and social well-being, justice, and resilience if the knowledge of how it develops does not remain in the hands of a few people with relatively marginal interests in such goals. It should instead become open source and available to anyone. Without transparency regarding who has created a given technology, based on what logic and aims, and which algorithms, codes, or considerations played a part in the development, interests cannot be aligned, or moral perspectives effectively contemplated. An open-source approach would also support communities with divergent needs and considerations to freely connect to pre-existing technologies and develop them further. Regarding all these preconditions, participation seems to be an immensely important factor as priorities cannot be defined nor interests aligned without this. Participation relates to the capacity of individuals to influence collective decisions and of stakeholders to channel their opinions and interests into community decision-making.

The final precondition identified in this research introduces a completely different dimension to the discourse: the reward and social recognition of activities that are useful to society. These activities cover undertakings beyond what is now considered paid work. As the vision says, in the future, people may have a real chance to be *"human beings rather than human doings"*. With entangled AI, they may have a chance to undertake meaningful activity. However, to reach this state, one prerequisite is to start rewarding the valuable contributions made by both humans and technology.

### 4.3 The most important instrumental variables in the system of entangled AI

Variables with at least three outward relationships encompass concepts that participants deemed instrumental in influencing the future of entangled AI. They are factors that either influence AI or where AI may have significant influence and, in some cases, simultaneously both. They are also features that enable the whole normative vision of the participants to emerge.

The quality of AI services and the prevalence of AI are inherently crucial variables in the system. It will be hard to talk about entangled AI if the level, breadth, diversity, and frequency of errors of the services provided by artificial intelligence do not improve over time, supporting the aims of the whole system. However, as long as quality improves, the beneficial impacts of AI can arise. The same holds for the prevalence of AI.

However, as the vision states, *"the existence of AI is not an end in itself"*, so quality and prevalence are only helpful as long as they serve higher aims. For example, one widely discussed means of achieving well-being and justice was considered to be the satisfaction of basic needs. AI can play a role in identifying these needs as well as creating, optimizing, distributing, and locating the resources necessary to fulfill them.

Most participants agreed that the domination of a profit orientation must subside if we are to move toward an ideal modus operandi and the more harmonious co-existence of the human, non-human, and technological world. A decrease in the proportion of human activities motivated by profit would mean that the beneficial impacts of AI can be more evenly and fairly distributed, and the fulfillment of needs can become more decentralized. Moreover, it would facilitate the fight against overconsumption, a fundamental hindrance to more efficient and just resource allocation.

### 4.4 Other variables that depict AI and its relation to humans

In the normative system of entangled AI, humans are autonomous being aware of the importance of collective action, aim to live in self-organized communities in harmony with the natural world, and are conscious of their mortality. Hence, entangled AI supports the emergence and/or the maintenance of these traits. Human autonomy and self-determination remain crucial for individual well-being, so artificial intelligence is not supposed to infringe upon the latter by creating psychologically harmful degrees of dependence on digital solutions or misusing personal data generated by the related systems. Individuals make informed decisions without coercion. This leads to the establishment of trust in AI as people have confidence that AI works to their advantage. People want to comprehend AI and have access to this knowledge in different depths and contexts. The solutions, motivations, circumstances, and stipulations associated with the operation of a given technology are transparent to society. Trust, knowledge, and transparency, directly and indirectly, lead to political support as citizens understand the need for policy interventions that, for example, help awareness surrounding AI or the diffusion of AI itself.

All interested individuals and communities have the opportunity to opt into or out of using AI. Access to AI is enabled for all, and a wide range of members of society do use artificial intelligence. This is further facilitated through the personalisation of MI solutions and services. AI generates an income and welfare surplus, which is fairly distributed. AI contributes to achieving collective goals and a high standard of public services through applications in healthcare, education, culture, and mobility. In terms of needs fulfillment, AI is also an optimisation tool that curbs the consumption of above-average needs or available resources, helping in the fight against overconsumption on the collective level and humans to stay within ecological boundaries.

However, AI is not without an ecological footprint, and throughout the development phase, it is clear that energy, raw materials, and other resources required to operate it is considered.

Decentralization was a crucial concept in participants' discourses. On the one hand, the strong tendency to advocate decentralization reflects the systems view of participants. They suggest that a centralized AI penetrating all spheres of life create significant dependencies and reduce the resilience of the system, even with precautions and ethical considerations. On the other, it reflects their quest for *"empowering shared power",* as participants phrased it in the vision. Human needs must be satisfied using diverse tools and approaches.

Arguments about power are reflected in the call for self-determination, meaning that individuals can influence their fate and exercise their rights on different social levels. They also presumed that rights can be better ensured in smaller communities based on subsidiarity and solidarity. Communities *"living in liquid democracies"* play an important role in this world of entangled AI. People self-organize according to values or interests. Social responsibility safeguards individuals' commitment to contribute to activities that serve the collective good. AI not only supports such concepts by providing platforms but also develops and operates according to these values.

A third perspective about decentralization emphasizes this with respect to AI ownership, claiming that without it, true resilience and shared power cannot exist.

## 5 Discussion

Our research aimed to reveal a normative future vision of AI technology that serves not just humanity but also the natural world and as an unintended discovery we adopted the concept of entangled AI. Applying Makridakis's (2017) typology (introduced in the first section of the paper), the majority of participants in the research can be classified as pragmatists. They were pragmatists in their approach about the necessity and power of regulation, and they also considered the inherent moral aspects in depth. To a small degree, they also showed signs of being doubters, leaving space for human creativity to flourish in the future while emphasizing human–AI collaborative creativity. Hence, participants could envision a livable future with compromises. The inherent methodological nature of the backcasting process helped avoid the generation of pessimistic and extreme outcomes and enabled participants to focus on a desired future. Even if this resulted in an idealistic approach, the method facilitated the consideration of a diverse range of issues, including basic needs, social well-being, moral balance, and sustainability. Such complexity in the final vision and the backcasting policy intervention steps can support critical thinking in policymaking.

In this participatory research, AI was envisioned to become the engine of human decisions, with the engine assumed to be also an "engineer" with hard and soft skills but only combined with a community of human experts capable of controlling the respective machines. This result is consistent with Laux's (2023) and Floridi's (2023) claim that human oversight is crucial for AI governance and that humans can support the future of AI proficiently and insightfully. However, human decisions do not reliably embrace non-human perspectives (such as that of the environment). Hence, AI must operate using more than human considerations if broader perspectives are to be included as fundamental parts of AI goals. The triangle of humanity, ecology, and technology indicates a more sophisticated and interconnected relationship.

Our key finding, the entanglement of AI assumes that AI should serve nature's purposes and support social well-being, justice, and the resilience of this intertwined network at the same time. The goal is not simply the existence of AI as a technology but the support for the dynamic balance of all these factors. Moreover, in their final round of reflections, participants noted how they had not discussed technology per se but how society, economy, ethics, and individual aspects are interrelated in critical ways in building an optimal future. Accordingly, not only the classic demands for AI, such as accountability and fairness, trustworthiness, transparency, and ethics (Jobin et al. 2019; James et al 2023; Larsson and Heintz 2020) featured prominently on the systems map but also resource demand, accessibility, decentralization, and inclusivity, employing a broader perspective. Thus, entangled AI assumes deep interconnection with socio-economic and governance processes which interconnectedness is also mentioned in Krüger's and Wilson's work (2022).

When assessing the results in terms of the ongoing scientific discussions, three perspectives are important. First, the research underpinned that reflecting on social factors (Feher and Katona 2021; Winiarska-Brodowska and Feher 2024), socio-sustainable issues (Geels and Smit 2000) and social responsibility is absolutely crucial in AI development—especially if we consider interconnectedness (Irwin and White 2019) at the individual as well as societal and environmental levels. Second, although some degree of technological determinism appeared in our findings too, the notion of human agency was strongly present. They were of the view that AI developments can trigger new human skills and change interests as well as collective human decisions (Beckert 2016; Dotson 2015), but in the envisioned adaptive society, humans consciously learn from technologically monitored feedback. Hence, the interconnectedness is not entirely determined by AI technology alone, as the vision allows for ethical choices, social diversification, and human deliberation in contrast to machine recommendations. Third, the results are controversial regarding the humanist–posthumanist discourse. On the one hand, AI was to promote the well-being of humans, relying strongly on human autonomy, consciousness, self-determination, and control over AI technology, and these point toward humanist elements in participants' thinking. At the same time, how interconnectedness emerged in the results strongly resonated with some aspects of posthumanism (Kalpokiene and Kalpokas 2023; Mellamphy 2021; Irwin and White 2019; Keeling and Lehman 2018).

Therefore, the findings of this backcasting research seem to transcend the dominant discourses related to expectations, technological determinism, and humanism. Even if humans still take center stage, technology can steward ecological sustainability as well as serve well-being. This entangled approach allows us to understand the pervasive role of technology, while permanent interdependence from technology is not an essential criterion. Instead, the elements in the triangle of nature–technology–humanity are mutually influential, and the level of dependence can change dynamically. Consequently, our research implies three distinct *policy implications* for entangled AI.

1. It is recommended that scientific debate mixes technological and non-technological concepts when discussing AI and related social aspects. Theories behind technological development may allow "out-of-the-box" thinking for the social sciences and humanities. Applying the term "entangled AI" confirms how beneficial this practice can be. Interdisciplinary terminology allowed us to reframe thinking about the impact of AI on the interconnectedness between technology, society, and nature. Technological determinism, the sociology of expectations, and other concepts need to be reconsidered and supported with new fields of meaning.
2. As a practical policy implication, it is highly recommended to apply the backcasting method to increase the understanding of interconnectedness and address blind spots in future planning. This verifiable method helps avoid a pessimistic approach by building a normative, pragmatist future vision. Additionally, changing values (Geels and Smit 2020) can also be interpreted through this process, circumventing path dependencies and lock-in effects of present mind frames and allowing the visualization in a broader field of interpretation.
3. For tech policy, it is recommended that experts' normative visions are applied from time to time in social thinking about AI futures. This type of research can be repeated regularly, and the results can be incorporated into policymaking. This can support proactive policy practice, combining technology-driven interconnectedness with human pragmatism. If pressure is significant, this proactivity may result in better functioning science–policy interfaces. Incentives and standards in policy decisions can be improved, and regulatory tasks can be reduced.

Overall, such backcasting research can provide fruitful visions and a more holistic view to policymakers whose goal is to support human autonomy and community norms in AI development. Moreover, the method is suited to connect popular approaches and action trends in the present, regardless of the accuracy of predictions about real-life developments. Hence, the complexity of AI development can be treated in an interconnected manner, using the systems view of entangled technology.

Entangled AI is a story about the harmony of AI–human symbiosis. It is a narrative about how humanity may coexist with a pervasive artificial intelligence in a way that results in greater life satisfaction through attention to well-being, justice, resilience, and care for the planet, and individual and/or community aspirations and the environment partly generated by artificial intelligence do not end up in irreconcilable conflict. This future is to be reached proactively by the help of regulation.

This direction has already started with the first significant regulation effort on AI, called AI Act. The EU Digital Strategy introduced this direction proactively years ago. The legislative framework addresses different rules for different risk levels (Laux 2023). Even if the AI Act needed a long preparation, the rising generative AI effect has accelerated the demand for the framework. Likewise, there is a drive to advance new standards through regulatory power in the United States. As per the Executive Order issued by the White House, the goal is a safe, secure, and trustworthy AI with a special focus on security and privacy.[1] The swift changes necessitate both proactive and reactive policy measures in the short terms. Thus, it remains a theoretical question of what would have happened if there had been an intensive focus on generative AI earlier and not after the technology-halting petition. Nevertheless, the rapid changes rather underline the need for proactive planning an optimal future, supported by the concept of entangled AI for policymakers. Our article has been dedicated to this policy support.

[1] https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

## 6 Conclusion

The purpose of our study was to introduce the original concept of entangled AI that emerged from participatory backcasting research on the future of AI. In their normative visions, the expert panel identified a liquid democratic system wherein AI mirrors the consciousness of human beings, supported by responsibility, societal harmony, solidarity principles, human autonomy, self-determination, deliberative ethical choices, and community norms. While the ultimate goal of AI remains to support individual well-being, this individualistic approach is not merely a selfish vision of the future. The projected vision emphasizes empowered and responsible individuals who safeguard the delicate balance between technology, humans, and nature. This way, humanity plays a vital role in shaping the ecological environment and staying in control of technology identifying optimal alternatives for individuals as well as collectives.

In light of the results that were obtained, the backcasting method is strongly recommended for normative analysis. This is due to its ability to build pragmatic future visions effectively. In addition, incorporating changing values and a dynamic level of dependence among the elements of the nature–technology–humanity triangle helps visualize a broader field of interpretation. It is therefore suggested that the backcasting method be used in parallel in forecasting research and modeling for policymaking.

Time constraints associated with the availability of participants did not allow all relevant aspects to be discussed. Thus, a few issues considered important components in the relevant literature and policy debates are missing from the research. For example, the dilemmas surrounding the well-known and controversial Chinese social credit score system were not mentioned. While problems related to demographic diversity, biases, inequalities, and cultural differences were mentioned, they were not considered in detail. Human augmentation, simulations, and virtual environments also remained generally undiscussed. All participants were European, but the EU regulatory environment was not even mentioned during the process. Another substantial limitation of the research is the homogeneity of the participants in terms of gender, race, and social and geographical background.

As for future research directions, exploring the potential of implementing a liquid democracy system that incorporates AI technology that reflects and addresses human consciousness, individual well-being, and basic needs could be a target for policymaking. This research could focus on the design of such a system, as well as the potential benefits and drawbacks of using entangled AI in democratic decision-making processes and its impact on ecological sustainability.

**Acknowledgements** The authors wish to thank the participants of the backcasting research for devoting their time to the project and sharing their insights with us. The authors also owe their gratitude to Róbert Pintér for his involvement in the research workshops as a facilitator. The authors thank Ágnes Horváth, Tamás Bokor, and Róbert Pintér for their help in the recruitment of the expert participants.

**Funding** Open access funding provided by Corvinus University of Budapest. The research was financed by the National Research, Development and Innovation Office of Hungary under the K131733 funding scheme. The contribution of Katalin Fehér was also supported by the Janos Bolyai Research Scholarship of the Hungarian Academy of Sciences under Grant No. BO/00045/19/9 and the European Union's Horizon Europe Research and Innovation Programme–NGI Enrichers, Next Generation Internet Transatlantic Fellowship Programme, Grant No. 101070125.

**Data availability** The method used in this research is of deliberative, qualitative nature. Hence, no quantitative datasets were generated or analyzed during the current study. The details of the research project, the methodology, and the outputs and results are published here: https://unipub.lib.uni-corvinus.hu/9810/. The voice recordings, written notes, and photos underpinning the outcomes of deliberations are available upon request.

## Declarations

**Conflict of interest** The authors have no financial or proprietary interests in any material discussed in this article.



## References

Andrada G, Clowes RW, Smart PR (2023) Varieties of transparency: exploring agency within AI systems. AI Soc 38(4):1321–1331. https://doi.org/10.1007/s00146-021-01326-6

Barad K (2007) Meeting the universe halfway: quantum physics and the entanglement of matter and meaning. Duke University Press, Durham

Barbrook-Johnson P, Penn AS (2022) Systems mapping: how to build and use causal models of systems. Springer Nature, Cham. https://doi.org/10.1007/978-3-031-01919-7

Beckert J (2016) Imagined futures: fictional expectations and capitalist dynamics. Harvard University Press, Cambridge, p 8825

Bergold J, Thomas S (2012) Participatory research methods: a methodological approach in motion. Histor Soc Res 37:191–222

Bignall S, Hemming S, Rigney D (2016) Three ecosophies for the anthropocene: environmental governance continental posthumanism and indigenous expressivism. Deleuze Stud 10(4):455–478. https://doi.org/10.3366/dls.2016.0239

Borup M, Brown N, Konrad K, Van Lente H (2006) The sociology of expectations in science and technology. Technol Anal Strateg

Manag 18(3–4):285–298. https://doi.org/10.1080/09537320600777002
Brown N, Michael M (2003) A sociology of expectations: retrospecting prospects and prospecting retrospects. Technol Anal Strateg Manag 15(1):3–18. https://doi.org/10.1080/0953732032000046024
Brown J (2010) The world café: shaping our futures through conversations that matter. ReadHowYouWant.com
Capra F, Luisi PL (2016) The systems view of life: a unifying vision. Cambridge University Press, Cambridge
Deleuze G, Guattari F (1987) A thousand plateaus: capitalism and schizophrenia (B. Massumi, Trans.). University of Minnesota Press, Minneapolis
Descola P (2014) Beyond nature and culture. The handbook of contemporary animism. Routledge, England, pp 77–91
Diakopoulos N, Friedler S, Arenas M, Barocas S, Hay M, Howe B, et al (2017) Principles for accountable algorithms and a social impact statement for algorithms. FAT/ML
Dotson T (2015) Technological determinism and permissionless innovation as technocratic governing mentalities: psychocultural barriers to the democratization of technology. Engag Sci Technol Soc. https://doi.org/10.17351/ests2015.009
Draude C, Klumbyte G, Lücking P, Treusch P (2020) Situated algorithms: a sociotechnical systemic approach to bias. Online Inf Rev 44(2):325–342. https://doi.org/10.1108/OIR-10-2018-0332
Feher K, Katona AI (2021) Fifteen shadows of socio-cultural AI: a systematic review and future perspectives. Futures 132:102817. https://doi.org/10.1016/j.futures.2021.102817
Feher K, Vicsek L, Deuze M (2024) Modeling AI trust for 2050: perspectives from media and info-communication experts. AI Soc. https://doi.org/10.1007/s00146-023-01827-6
Floridi L (2023) AI as agency without intelligence: on ChatGPT, large language models, and other generative models. Philos Technol 36(1):15. https://doi.org/10.1007/s13347-023-00621-y
Geels FW, Smit WA (2000). Potholes in the road to the future: Lessons and pitfalls from failed technology futures. In: Contested futures: a sociology of prospective techno-science. Ashgate. pp. 129–155
Glenn JC (2009) The futures wheel. Futures research methodology—version 3 19
Gordon G (2023) Digital sovereignty, digital infrastructures, and quantum horizons. AI Soc. https://doi.org/10.1007/s00146-023-01729-7
Haimes YY (2002) On the complex interdependencies of infrastructure systems. J Urban Technol 9(1):109–123
Herbrechter S, Callus I, de Bruin-Molé M, Grech M, Müller CJ, Rossini M (2022) Critical posthumanism: An overview. In: Palgrave handbook of critical posthumanism. pp. 3–26
Herzog C, Diebel-Fischer B (2022) Mapping the landscape of sociotechnical entanglement. Ethics Inf Technol 24(2):99–114
Irwin R, White TH (2019) Decolonising technological futures: a dialogical tryptich between Te Haumoana White, Ruth Irwin, and Tegmark's artificial intelligence. Futures 112:102431. https://doi.org/10.1016/j.futures.2019.06.003
James A, Hynes D, Whelan A, Dreher T, Humphry J (2023) From access and transparency to refusal: three responses to algorithmic governance. Internet Policy Review. https://doi.org/10.14763/2023.2.1691
Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nat Mach Intell 1(9):389–399. https://doi.org/10.1038/s42256-019-0088-2
Kalpokiene J, Kalpokas I (2023) Creative encounters of a posthuman kind–anthropocentric law, artificial intelligence, and art. Technol Soc 72:102197. https://doi.org/10.1016/j.techsoc.2023.102197
Kautz KK, Blegind JT (2012) Debating sociomateriality: entanglements, imbrications, disentangling, and agential cuts. Scand J Inf Syst 24(2):5
Keeling DM, Lehman MN (2018) Posthumanism. Oxford Res Encycl Commun. https://doi.org/10.1093/acrefore/9780190228613.013.627
Kerr A, Barry M, Kelleher JD (2020) Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance. Big Data Soc. https://doi.org/10.1177/2053951720915939
Konrad K, Van Lente H, Groves C, Selin C (2016) Performing and governing the future in science and technology. Handb Sci Technol Studi 465–493
Köves A, Vicsek, LM, Fehér K, Fischer M (2023). Artificial intelligence serving the future. Results of the backcasting workshop with experts. https://unipub.lib.uni-corvinus.hu/9810/ Accessed Aug 8, 2024
Köves A, Király G (2021) Inner drives: Is the future of marketing communications more sustainable when using backcasting? Futures 130:102755. https://doi.org/10.1016/j.futures.2021.102755
Kriman AI (2019) The idea of the posthuman: a comparative analysis of transhumanism and posthumanism. Russ J Philos Sci. https://doi.org/10.30727/0235-1188-2019-62-4-132-147-132-147
Krüger S, Wilson C (2023) The problem with trust: on the discursive commodification of trust in AI. AI Soc 38(4):1753–1761
Kubes T, Reinhardt T (2022) Techno-species in the becoming towards a relational ontology of multi-species assemblages (ROMA). Nano Ethics. https://doi.org/10.1007/s11569-021-00401-y
Larsson S, Heintz F (2020) Transparency in artificial intelligence. Internet Policy Rev. https://doi.org/10.14763/2020.2.1469
Laux J (2023) Institutionalised distrust and human oversight of artificial intelligence: toward a democratic design of AI governance under the European Union AI Act. Available at SSRN 4377481
Lee J (2023) The era of omni-learning: frameworks and practices of the expanded human resource development. Organ Dyn 52(1):100916. https://doi.org/10.1016/j.orgdyn.2022.100916
MacKenzie D, Wajcman J (1999) The social shaping of technology. Open University Press, Berkshire
Makridakis S (2017) The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. Futures 90:46–60. https://doi.org/10.1016/j.futures.2017.03.006
Mellamphy NB (2021) Humans "in the Loop"? Human-centrism, posthumanism, and AI. Nat Cult 16(1):11–27. https://doi.org/10.3167/nc.2020.160102
Permin E, Bertelsmeier F, Blum M, Bützler J, Haag S, Kuz S et al (2016) Self-optimizing production systems. Proc Cirp 41:417–422. https://doi.org/10.1016/j.procir.2015.12.114
Powell AB, Ustek-Spilda F, Lehuedé S, Shklovski I (2022) Addressing ethical gaps in 'technology for good': foregrounding care and capabilities. Big Data Soc. https://doi.org/10.1177/20539517221113774
Putman VL, Paulus PB (2009) Brainstorming, brainstorming rules and decision making. J Creat Behav 43(1):29–40. https://doi.org/10.1002/j.2162-6057.2009.tb01304.x
Quist J, Vergragt P (2006) Past and future of backcasting: the shift to stakeholder participation and a proposal for a methodological framework. Futures 38(9):1027–1045. https://doi.org/10.1016/j.futures.2006.02.010
Quist J, Thissen W, Vergragt PJ (2011) The impact and spin-off of participatory backcasting: From vision to niche. Technol Forecast Soc Chang 78(5):883–897. https://doi.org/10.1016/j.techfore.2011.01.011
Robinson J (2003) Future subjunctive: backcasting as social learning. Futures 35(8):839–856. https://doi.org/10.1016/S0016-3287(03)00039-9

Rovelli C (2022) Helgoland: making sense of the quantum revolution. Penguin. ISBN10–0593328892

Timmermans S, Tavory I (2012) Theory construction in qualitative research: from grounded theory to abductive analysis. Sociol Theory 30(3):167–186. https://doi.org/10.1177/0735275112457914

Van Lente H (2012) Navigating foresight in a sea of expectations: lessons from the sociology of expectations. Technol Anal Strateg Manag 24(8):769–782. https://doi.org/10.1080/09537325.2012.715478

Van Noordt C, Misuraca G (2022) Exploratory insights on artificial intelligence for government in Europe. Soc Sci Comput Rev 40(2):426–444. https://doi.org/10.1177/0894439320980449

Vicsek L (2021) Artificial intelligence and the future of work–lessons from the sociology of expectations. Int J Sociol Soc Policy 41(7/8):842–861. https://doi.org/10.1108/IJSSP-05-2020-0174

Winiarska-Brodowska M, Feher K (2024) Socio-cultural artificial intelligence (SCAI) and journalism: a transdisciplinary perspective in media and communication studies. In: Brzeziński D, Filipek K, Piwowar K, Winiarska-Brodowska M (eds), Algorithms, artificial intelligence and beyond. Theorising Society and Culture of the 21st Century, Routledge

Zou J, Schiebinger L (2018) AI can be sexist and racist—it's time to make it fair. Nature. https://doi.org/10.1038/d41586-018-05707-8