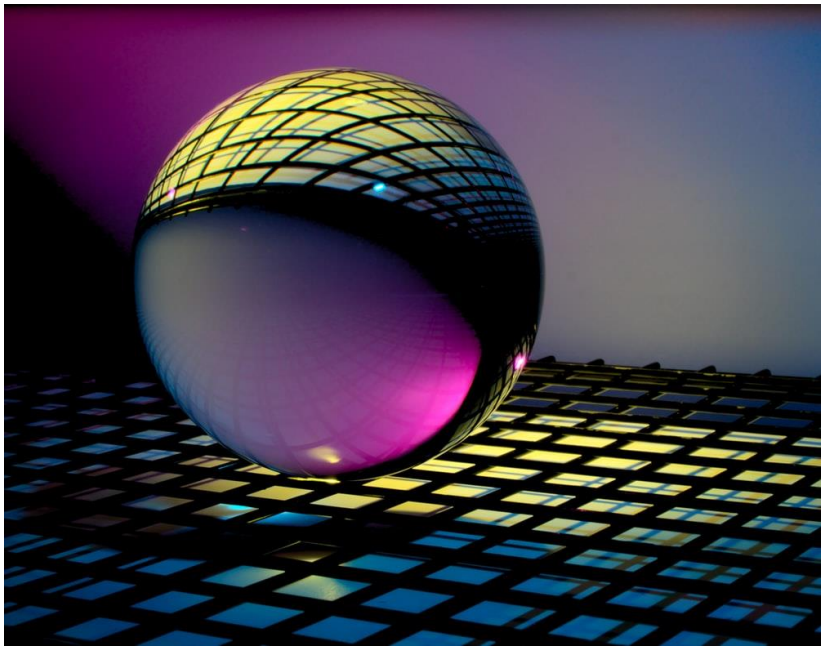
CORVINUS UNIVERSITY of BUDAPEST

# RESEARCH REPORT

## ARTIFICIAL INTELLIGENCE SERVING THE FUTURE – Results of the backcasting workshop with experts

## Budapest, 2023



**Alexandra Köves**

**Lilla Vicsek**

**Katalin Fehér**

**Máté Fischer**

# Content

## Research background

The workshop took place in the framework of the OTKA research K-131733 "*Visions on artificial intelligence and society*", funded by the National Research, Development and Innovation Office in Hungary. The research focuses on visions of AI and their impact. On the one hand, it analyses dominant visions of AI, their implications and shows how dominant visions contribute to closing off certain possible future alternatives. On the other hand, it explores alternative desirable futures.

The research rejects the technological deterministic view of technological development, which views technology as a phenomenon external to society and in which society's only task is to adapt to technological changes. Social science research demonstrates that it is more useful to conceive change in terms of interaction between social and technological factors. It is therefore important to consider and make it explicit how social factors influence the development of technologies. From this perspective, the research also explores how technological developments could help to create a desirable future state. To this end, it is important to clarify what is considered a desirable state by whom and how such a state can be achieved.

To answer these questions, it is important to break away from the traditional assumptions that determine our current thinking, and to find innovative ways to do so. For this purpose, we have chosen the method of backcasting.

## The aim of the workshop

The aim of the workshop was to find out what AI experts would like to see in the future for AI for 2060, and to explore what steps they would consider important to make such a future reality.

The title of the workshop was "Artificial Intelligence serving the future". One of the elements of the title was "Artificial Intelligence". There are many definitions of the term. Whole workshops could be held on how to define this term. In our research, we use the definition of Tegmark (2017) as a starting point to define artificial intelligence as the ability of a machine, computer, or software to achieve complex goals. The approach assumes that machine learning is part of AI. Another element of the title is "serving the future". Here we have deliberately chosen this broad formulation, including considerations for the environment in addition to serving humanity.

## Research methodology: backcasting

Today, we are confronted with a number of complex, systemic problems that are both difficult to address within existing paradigms and where atomistic approaches, i.e. breaking the problem down into parts, can lead to a loss of understanding and identification of appropriate points of intervention. Backcasting is a method of normative scenario building that explores the possible futures of a social organisation (be it a company, a city or even a whole society). The novelty of the method lies in the fact that, rather than using the current situation as a starting point for predicting an expected future, it starts by constructing an ideal normative vision of the future, and links this vision - going backwards in time - to the present, identifying the steps that could lead to the imagined future (Quist and Vergragt, 2006; Robinson, 2003). The English name of the technique refers to this when it replaces the prefix 'fore/casting' with 'back/casting'.

The method assumes that our current mindset and the closing effects of the present can adversely affect the actions we are willing to take now. Consequently, the actors' vision of the future also has a strong influence on our actions. In this, it differs significantly from the basic principles of foresight. Another important difference is the perception of actors. While foresight assumes that actors simply drift with external trends and events, backcasting relies on the assumption that actors themselves can move towards certain futures, and in feedback processes, action and visions interact incessantly, so that both actions and future directions can be influenced.

Backcasting is particularly effective in environments where the future is uncertain, the heterogeneous systems involved are complex, and the outcomes that can be predicted from current trends are unacceptable (Quist and Vergragt, 2006). Backcasting is therefore most necessary when systemic innovation is essential to break away from current trends. In sum, this normative approach has the potential to generate alternative visions of the future that go beyond mainstream frameworks over a longer time horizon and across a wide range of issues, as well as to identify innovative intervention opportunities that may not yet be present or sufficiently emphasised in current thinking. The role of artificial intelligence in the service of the future is a highly complex problem to which the chosen methodology is well suited.

## The research process

The research involved 15 experts from diverse backgrounds: AI developers from multinational companies from the automobile, electric and energy industries; AI experts from the consulting and legal sector; a journalist covering AI development topics; leading academics with diverse technology-oriented topics, representatives of NGOs with technology related missions and two PhD students doing research on AI.

Our research employed participatory backcasting. The concept of participation is founded on the assumption that participants can form views and make decisions about matters that are important to them, even if they lack in-depth expertise either in the given field or concerning specific perspectives associated with the deliberations (Bergold and Thomas, 2012). In a participatory backcasting process future scenarios are not created by researchers, but by the participatory process itself. In these instances, future visions at the heart of the analysis are defined based on the preferences of participants instead of external factors and expectations (Robinson, 2011). According to Quist and Vergragt (2006), there are five main steps in participatory backcasting: (1) Strategic problem orientation; (2) Future vision development; (3) Backcasting analysis; (4) Future alternative – follow-up agenda elaboration; and (5) Follow-up. The last step, follow-up action, has not yet taken place within the current backcasting project; hence will not be detailed here.

The backcasting process consisted of four main phases that fit well with the steps identified above. Although participatory backcasting does not require participants to be experts in the given field, in this case, we chose for a panel the members of which were somewhat connected to artificial intelligence in particular or technology in a broader sense through their work. Despite our best efforts, we were only able to obtain diversity in terms of sector; nonetheless, the project did not meet the requirements for gender and race diversity, as all of the participants—all except one—were White men. Regrettably, this circumstance illustrates the range of individuals working on AI development.

| Backcasting stage (based on Quist et al., 2011) | Research Methodology | Outputs |
|---|---|---|
| 1. Strategic problem orientation | Brainstorming with workshop techniques Putman ,Paulus[40] | Future vision fragments |
| 2. Future vision development | World Café Brown, Issaacs[41] | The future vision itself |
| 3. Backcasting analysis | Futures Wheel Glenn [42] – slightly modified | Identified backcasting steps (tools and recommendations for intervention) |

| | | |
|---|---|---|
| 4. Future alternatives | Systems mapping<br>with the casual loop diagram technique Barbrook-Johnson, Penn [17]) | Key variables and relationships between them |

The process of our backcasting research

On the first day of the workshop (20th of May 2022), backcasting participants defined how artificial intelligence may be able to shape the future. Due to the limited time available, specific themes were predefined, facilitating the beginning of collaborative thinking. The four main topics selected as starting points for addressing the broader issue were:

(1) AI supporting the satisfaction of human needs;

(2) AI supporting fair work and livelihoods;

(3) AI supporting environmental regeneration

(4) AI supporting individual and collective decision-making.

The topics served for the purpose of orientation but were not intended to limit the framework of thinking since the goal was precisely to enable viewpoints that would help move away from the lock-in effects of the present.

First, the participants were prompted to create ideal visions of the future relying on headlines based on the previously defined themes. Then, the vision fragments that emerged were grouped by the researchers around five major characteristics: (1) Sustainable and regenerative; (2) Fair and liveable; (3) Self-transcending; (4) People- and life-oriented; and (5) Exploratory. Together with the four pre-defined themes, these represented the foundation of the shared framework of thought for the subsequent phases.

The World Café approach (Brown and Isaacs, 2005) was the methodology for future vision development. Here, participants were given guidance to think associatively and creatively while focusing on elaborating innovative ideas instead of analytical problem-solving. The pre-determined topics were discussed at four different tables, with groups of participants rotating between them, enabling everyone to express their thoughts about each topic. The researchers took on the role of facilitators at the tables and assisted participants in the process of developing visions while taking care to keep the focal points clear without influencing their ideas and ensuring that their viewpoints were fairly represented in the conversation. During the World Café, the most important thoughts were noted, and audio recorders also helped with the documentation. Relying on all these methods, the researchers then created a shared vision by synthesizing ideas that arose at different tables. Participants evaluated the vision before the second workshop, and at the beginning of the second workshop, the final form of the text was created.

To achieve the desired future state embodied by the vision, on the second day of the workshop (3rd of June 2022) the participants determined four key areas of action that can be seen in the table below. To determine the specific steps related to these, a modified version of the Futures Wheel method (Glenn, 2009) was utilized. When working with a Futures Wheel, participants create a chain of future consequences in terms of the investigated phenomenon, in the process constructing a multi-layered network of implications. In the case of the current research, the operational logic of the methodology was reversed; the participants defined the steps as prerequisites by moving backwards in time from the created future vision instead of focusing on the future consequences of a present phenomenon. In practice, during the second workshop, participants in small groups defined goals related to the four key areas (at least three in each case) for 2060. After this, the intervention steps that would support the achievement of the final goals were first defined for 2050, then for 2040, and finally for 2025. Finally, the groups presented their backcasting roadmaps and discussed the synergies and contradictions arising regarding their respective reversed Futures Wheels.

| Key areas of actions | Description |
|---|---|
| 1. Creation of decentralized artificial intelligence | In the future, social decisions and the use and ownership of artificial intelligence must be decentralized, as this can best support the diversity, resilience, and justice of systems. |
| 2. Ensuring the positive impact of artificial intelligence on social well-being | AI increases efficiency and generates significant revenue, enabling a wide range of basic needs to be fulfilled, but only if we design a fair distribution system. |
| 3. Creation of reliable and transparent artificial intelligence for decision support | Artificial intelligence can serve humanity as a form of decision support for handling complex issues if people trust it. However, this requires that ethical considerations are taken into account, that AI is transparent to at least certain groups, and that it operates with the smallest margin of error. |
| 4. Creation of harmonious coexistence of artificial intelligence with human and ecological environment | Artificial intelligence can only support the future of both the ecological environment and humanity if it achieves harmony with both in its operations. |

Key areas of action for achieving the desired future state

Finally, due to time constraints, the researchers compiled a logical arrangement of the roadmap items. The researchers chose to display the logic of the participants' arguments in a systems map in order to highlight the linkages that surfaced behind the components of the visions, the goals, and the intervention steps (Barbrook-Johnson and Penn, 2022). Systems maps serve to depict the combination and relationships of elements that make up an entire system. Therefore, they often feature networks of nodes and edges that indicate the type and strength of relationships between the components. Among the many types of systems mapping methodologies, a slightly modified version of a causal loop diagram technique was constructed. The researchers examined the future vision's text, as well as the objectives and suggested intervention measures put forth by the participants, in order to identify the variables of the systems map and systematize the backcasting results. The mapping procedure followed next, when participants' arguments were used as the source to show the linkages between the categorized variables and the reinforcing (positive) and balancing (negative) feedback loops. The colour-coded categorization in the chart indicates whether an element is a prerequisite/target/instrumental variable or the fundamental aim itself. Furthermore, a table describing all the variables of the system map can also be found under the results section.

# Results

This research report contains the direct results of the process and does not include any evaluation or analysis. In addition to the systems map, the results presented are a synthesis of what was said by the participants, edited into a single version by the researchers.

## The vision

During the backcasting workshop, participants created the following vision (translated to English):

*In 2060, artificial intelligence will act as an engine for human decisions in order to enable better choices, especially on complex issues. We regard artificial intelligence as part of our environment, like the ecological environment, and strive to live in a harmonious relationship with both. Artificial intelligence is like electricity used to be. 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. In the case of professional decisions, AI metaphorically provides a pair of glasses for the decisions. In many cases, AI makes decisions without human intervention, but in self-organising groups, people still take care that these decisions are ethically sound. Alongside ongoing value choices, there are discourses about ensuring that decisions by AI can be overridden. Business co-operatives operate in virtual meta-spaces, with transparency guardians providing oversight (similar to the UN or EU operating in the twenties). Decisions are primarily based on long-term goals, and this [process] is supported by AI. In addition to business logic, other kinds of decision-support rationalities also appear.*

*A large part of humanity lives in liquid democracies, which means the widest possible participation, transparency of stakeholders' perspectives, decentralisation of decisions, with technological interoperability, in which individual input, self-organisation and transparency are key. All of this can be supported by AI, while at the same time providing an overview of the impact of decisions on other groups or on the living and non-living environment. In this way, MI also plays a certain coordinating role, similar to the role of the state, but it builds on broader information, fairness and ethics. Corruption has thus been almost completely eliminated. Joint decisions are supported by the networks of contacts and everything happens in joint interactions with artificial intelligence. However, this environment does not mean that everyone is involved in socio-economic decisions. For those who are unwilling or unable to participate, the option to opt out is open. However, AI can channel and contextualise individual situations in participatory democracies. AI promotes decentralised power sharing (empowering shared power) and also seeks to avoid the tyranny of the majority in collective decision-making. AI makes political content visible. Optimisation is selective: there are things to optimise for (e.g. fewer accidents), but there are still things not to optimise for. There is a moral-emotional-instinctual level in both individual and collective decisions. Society in 2060 is still a society influenced by people's free will.*

*Despite the widespread use of artificial intelligence, people in 2060 will be free to make their own choices and take responsibility for their decisions. However, these decisions will be supported by AI as an assistant, tailored to the individual. Privacy is fully guaranteed. The possibility of human spontaneity and human irrationality is preserved. However, AI is able to support any decision, so it can advise people by maximising the happiness index. For example, if a person wants to move, AI can suggest which neighbourhood they and their family would best fit into, or AI can help design a tailored diet to recommend the most optimal combination of health and pleasure for that individual. At the same time, the individual retains the freedom to make decisions without the advice of the AI, even if this would be to their own detriment. If one wants to become an alcoholic, s/he can turn off the technology that warns of the negative consequences of the actions. While our lives are largely permeated by AI, there are areas (e.g. human relationships, empathy, emotions, physical contact, personal creative activities, some aspects of education) that are not dominated by technology.*

*In 2060, work is not just employment in the workplace, but value-creating activities in communities. It is an activity that is personalised and in harmony with nature. Work is not done for a living and livelihoods are not dependent on activity in the labour market. Decent work, defined in 2060 rather as meaningful activity, and the provision of basic security needs are the primary goals supported by AI as a tool. Thus, the existence of AI is not an end in itself, but, among other things, a means to help with these. It allows everyone to do what they want to do - even if, for example, AI is more effective at it. You still see people driving trucks on the roads, but it is no longer a matter of coercion but of choice. Meaningful activity can be not just work, but any activity, cause, project, creative pursuit, art, even play. The system is able to handle the diversity of people: some people want to work in one place for a long time; some people want to change jobs more than once; some people want to work in a creatively self-organising way; and it is accepted that some people do not want to work.*

*People do what they love, according to their own sense of purpose, mission and inner motivation. In this way, AI supports people's individual development. AI can do the jobs that people don't like (for example, boring, repetitive, dirty work), but there are still tasks - mostly value-added, human-interaction activities - that people prefer to do. However, people in 2060 are defined not by what they do, but by how they exist. Thus the term "human being" rather than "human doing" makes sense again. As the classic workplace environment disappears, the strongest communities will be built around interests, and AI will help to find them.*

*Everyone is entitled to certain basic services and basic quality of life. In 2060, basic needs are also basic rights. It is fair because everyone can meet their needs and is not forced to do something they don't want to do. But equity does not mean that everyone gets the same benefits/basic income, differences and diversity in needs are taken into account. A high level of personalised, experiential education is available to all. Together, basic income/benefits and technological progress create the conditions for us to live equitably and not be driven by coercion. At the same time, people can enter into the working to satisfy their on needs at any point. If they want to garden and grow their own food, they can do so. At the same time, AI not only supports the development of one's skills, but also helps to bring to the surface needs that have been suppressed in humans, for example in the arts. In addition to equity, AI also has a major role to play in supporting health. Thanks to it, 100-year-old people are cycling on the streets. Disease disappears, but death does not. But society can prepare for death, and the journey is not only longer but also more informed.*

*The value created by the AI becomes part of public services, so that AI can address the historical problems of redistribution in a fair and impartial way. The scarcity of access due to environmental conditions (e.g. resource constraints) and/or the basic nature of an activity (e.g. going to the theatre or travelling abroad) can be addressed by data-based AI in a way that is both accessible to all and transparent about how it is distributed. Even in 2060, you might need to wait for months to be able to go to a theatre performance, although the full interoperability of real and virtual space can reduce these demands. However, in 2060, we are not a society based on pleasure and growth, but on value creation, and so the full satisfaction of needs beyond necessity is no longer a priority.*

*In this value-based society, by 2060, the partnership between AI and humanity has made significant progress in regenerating our natural environment. Instead of the previous approach of lording over nature, we are striving for harmonious coexistence and a state of balance, and our consumption patterns have been transformed in line with this. AI-based incentive schemes to achieve optimal individual consumption levels and ecological footprints, while protecting individual rights, can ensure that we do not over-consume.*

*With AI, energy production based on renewables and fusion energy is fully automated and the associated pollution is minimal. With the support of AI, we have created decentralised energy grids covering large geographical areas, where energy production and consumption are often realized at a single point: for example, smart homes or cars. Biodiversity is being restored and there are no longer islands of rubbish in the oceans. We are surrounded by advanced aquaculture. Sophisticated product life-cycle analyses, logistics systems and biodegradable materials with programmable properties, such as biopolymers, supported by AI, contribute greatly to the predictability and optimal functioning of waste management. Through AI and a circular approach, global material and energy flows are efficient. There is no food waste. In 2060, our cities have self-driving cars with AI control that can be easily and quickly accessed by the masses. Our living space systems will be designed with the preservation of our natural environment in mind. So in 2060, our cities will not be dominated by concrete alien structures, but rather by green surfaces and the experience of being embedded in nature. Achieving harmony with our surroundings has made it possible to increase the vertical and horizontal extent of our cities, reducing population density and the sense of congestion.*

*We see technology and AI as an important tool, but only as a tool, for solving sustainability problems that may arise, whether ecological or social. Non-technological solutions are equally important, and we only use these technologies once we have understood the problem at the conceptual level, made the right value choices and taken into account the necessary moral considerations. We, humans.*

*Technology and AI are not helping us to avoid falling into the abyss or to stay just within the limits of the Earth's carrying capacity. In 2060, it is no longer a question of the survival of human societies, but of the extent to which they prosper.*

*AI is transparent, explainable, well documented, traceable, its operation and maintainance democratic. It is self-evolving, while human understanding, regulation and ethical frameworks evolve with it. Its operation and ownership is decentralised, not hierarchically organised, and does not support the dominance of groups of people. AI algorithms are able to maintain the most diverse solutions possible for resilience (e.g. what energy it uses, what values it supports, how to exit its systems). 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 incorporation of ethicality into AI also helps to overcome fears of singularity. The emphasis is on trust-based, transparent, conscious relationships between humans and AI, but with the possibility of total withdrawal from these relationships at both individual and community levels. At the same time, AI is intimately connected to the physical world, and thus has the capacity to feed back when we go beyond the physical resources available. This also helps to ensure that humanity does not become totally dependent on technology and that the conditions for life are maintained in the event of disruption (e.g. solar flares).*

## Intervention steps linked to the vision

On the basis of the vision developed during the first phase of the process, the participants selected four focal points for the second phase in order to develop recommendations for achieving the desired state. These four foci were as follows:

1) **Decentralisation of AI:** The initial argument was that in the future, social decisions and the use and ownership of AI should be decentralised, as this would best serve the diversity, resilience and fairness of systems.

2) **Ensuring the positive impact of AI on social well-being**: the initial argument was that because AI increases efficiency, it generates high revenues, thus enabling basic needs to be met, provided revenues are properly distributed. However, this equitable distribution must be established.

3) **Creating trustworthy and transparent AI for decision-support**: The initial argument was that AI can serve humanity as a decision-support system for complex issues if people trust it. However, this requires that its operation a) follow ethical considerations; b) be transparent at least to certain groups; c) be within the smallest possible margin of error.

4) **Creating a harmonious co-existence of AI with the human and ecological environment**: the initial argument was that AI can only serve the future of both the ecological environment and humanity if it works in harmony.

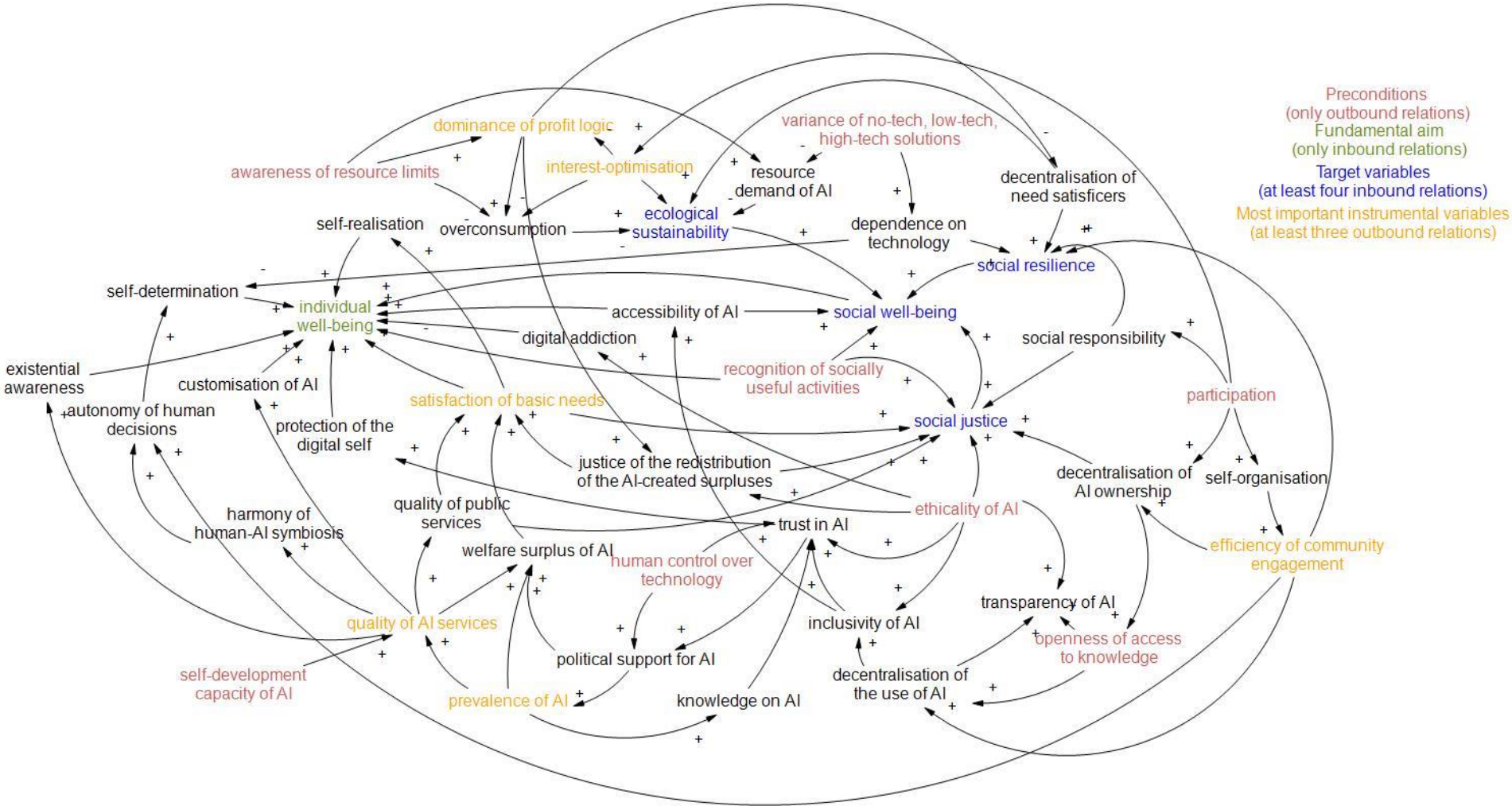
The first of these four focus areas, their associated vision elements, objectives, and recommended actions are set out in the table below. The table also indicates whether the steps include positive and/or negative regulatory, economic incentive-based or information/attitudinal interventions, and to which variable in the systems map they are most closely related. This is followed by a description of the systems map and its elements.

| Central theme | Target | Intervention step | Timeframe | Related item on the taxonomy map | Instrument: positive regulation | Instrument: negative regulation | Instrument: positive economic incentives | Instrument: negative economic incentives | Instrument: positive information | Instrument: negative information |
|---|---|---|---|---|---|---|---|---|---|---|
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Introduction of opt-out systems | 2050 | autonomy of human choices | x | x | x | x | x | x |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Setting rules on access to data | 2040 | digital self-protection | x | x | | | | |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Enabling certain social functions to be organised without AI | 2040 | social resilience | x | | x | | x | |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Providing opt-out options for sharing information about ourselves | 2025 | autonomy of human decisions | x | x | x | x | x | x |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Meeting local needs | 2050 | decentralisation of the provision of needs | x | | x | x | x | x |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Promoting local value creation instead of resource centralisation and redistribution | 2040 | Decentralisation of the use of AI | x | | x | | x | |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | 1-piece flow, launching pilots of value creation based on a string principle | 2025 | decentralisation of the provision of needs | | | x | | | |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Creating the conditions for hybrid ownership | 2040 | Decentralisation of ownership of AI | x | x | x | x | x | x |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Creating the legal conditions for community/user ownership | 2025 | Decentralisation of ownership of AI | x | | | | | |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Balancing and pluralising databases (Development and expansion of databases to include excluded groups) | 2025 | Inclusiveness of AI | x | x | x | x | x | x |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Reimbursement of the cost of data inserted | 2025 | Fairness in the distribution of benefits from AI | x | x | x | x | | |
| Decentralisation | Eliminating opportunities that would give control to the few over the many | Technological realisation of full digital identity ownership | 2050 | digital self-protection | x | x | x | x | x | x |
| Decentralisation | Eliminating opportunities that would give control to the few over the many | Ensuring full control and autonomy of digital identity | 2040 | digital self-protection | x | x | x | x | x | x |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Achieving full digital identity ownership | 2025 | digital self-protection | x | | | | | | |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Community ownership of databases of individual data | 2025 | Decentralisation of ownership of AI | x | x | x | x | | | |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | Only non-profit organisations above a certain number of employees can create a personal profile | 2040 | dominance of profit considerations | | x | | | | | |
| Decentralisation | Eliminate opportunities that would give control to the few over the many | For-profit organisations can only create a personal profile with personal consent | 2025 | digital self-protection | | x | | | x | x | |
| Decentralisation | Make decentralised resources more efficient than centralised | Transforming consumer behaviour towards sufficiency | 2050 | overconsumption | x | x | x | x | x | x | |
| Decentralisation | Make decentralised resources more efficient than centralised | Transforming lifestyles for quality of life | 2025 | individual well-being | x | x | x | x | x | x | |
| Decentralisation | Make decentralised resources more efficient than centralised | Creating the infrastructure for decentralisation (e.g. energy) | 2050 | Decentralisation of the use of AI | x | x | x | x | x | x | |
| Decentralisation | Make decentralised resources more efficient than centralised | Making our know-how openly available | 2040 | open access to knowledge | x | x | x | x | x | x | |
| Decentralisation | Make decentralised resources more efficient than centralised | Developing business solutions for open access to knowledge | 2025 | open access to knowledge | x | x | x | x | x | x | |
| Decentralisation | Make decentralised resources more efficient than centralised ones | Consciously extending the Creative commons system | 2025 | open access to knowledge | x | x | x | x | x | x | |
| Decentralisation | Building a liquid democracy | Strengthening individual autonomy by strengthening effective community functioning | 2050 | effectiveness of Community operations | x | x | x | x | x | x | |
| Decentralisation | Building a liquid democracy | Generalising experiential education | 2040 | quality of Community services | x | | x | | x | | |
| Decentralisation | Building a liquid democracy | Teaching artificial intelligence to think in a participatory way | 2040 | Participation | x | x | x | x | x | x | |
| Decentralisation | Building a liquid democracy | Making technology choices a substantive issue for participatory democracy | 2040 | Participation | x | x | x | x | x | x | |
| Decentralisation | Building a liquid democracy | Launching a social dialogue with test boards | 2025 | Participation | x | x | x | x | x | x | |
| Decentralisation | Building a liquid democracy | Individual and group development of value-creating skills | 2025 | dominance of profit considerations | x | x | x | x | x | x | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Decentralisation | Building a liquid democracy | Designing an experiential education system | 2025 | quality of Community services | x | x | x | x | x | x |
| Decentralisation | Building a liquid democracy | Developing business models for value creation | 2025 | dominance of profit considerations | x | x | x | x | x | x |

# The systems map



| VARIABLES IN THE SYSTEMS MAP | DEFINITION |
|---|---|
| **PRECONDITIONS FOR THE SYSTEM TO WORK (OUTPUT CONNECTION ONLY)** | |
| **AWARENESS OF RESOURCE LIMITS** | Acceptance and better understanding of the extent to which there are physical limits to the resources needed for human activities, both to respect the Earth's carrying capacity and to ensure a life-sustaining environment for people. |
| **ETHICALITY OF AI** | The acceptance and integration into the development and operation of AI that the operation of AI requires broad ethical consideration. |
| **SELF-DEVELOPMENT CAPACITY OF AI** | The extent to which the AI will be able to do more than it can now, and how the pace of its development will change. This includes the extent to which it will be able to develop itself without human intervention. |
| **VARIANCE OF NO-TECH, LOW-TECH, HIGH TECH SOLUTIONS** | The extent to which solutions that do not require technological intervention, those that rely only on basic tool requirements, and those that require complex technology coexist, complement each other and maintain harmonious relationships. The same problem should be addressed by building on as wide a range of these as possible. |
| **PARTICIPATION** | The extent to which individuals can influence collective decisions and the extent to which stakeholders can channel their own opinions and interests into community decision-making. |
| **RECOGNITION OF SOCIALLY USEFUL ACTIVITIES** | The extent to which activities (not just paid work in the current sense) that are useful to society are rewarded and socially recognised. |
| **HUMAN CONTROL OVER TECHNOLOGY** | The extent to which people are able to see and influence the development paths and guiding principles of AI, e.g. through communities of experts. |
| **OPEN ACCESS TO KNOWLEDGE** | Transparency of who/what logic, what algorithms, what code, what considerations have been used to create a technology, and how freely others can connect to it and develop it. |
| **FUNDAMENTAL AIM OF THE SYSTEM (INPUT ONLY)** | |
| **INDIVIDUAL WELL-BEING** | Subjective assessment of individual quality of life. A measure of how well an individual feels in life, going beyond the dimensions of material well-being. |
| **KEY TARGET VARIABLES OF THE SYSTEM (AT LEAST 4 INPUT CONTACTS)** | |
| **ECOLOGICAL SUSTAINABILITY** | The success of efforts to maintain and improve the ecological state of our planet. |
| **SOCIAL JUSTICE** | The fairness of the distribution of goods, opportunities and duties among members of society, taking into account the ethical norms of the community. |
| **SOCIAL WELL-BEING** | The health of the relationships and systems that bind individuals together; the alignment of people's social and individual roles. |
| **SOCIAL RESILIENCE** | The ability of a society to take over the functions of certain elements of the system and restore the equilibrium situation when the equilibrium of certain elements of the system is upset by external or internal influences. |

| KEY INSTRUMENTAL VARIABLES OF THE SYSTEM (AT LEAST 3 OUTPUT CONNECTIONS) | |
|---|---|
| **SATISFACTION OF BASIC NEEDS** | The extent to which the resources necessary to meet basic needs (e.g. housing, drinking water, food, health, education) are available to all, and to what extent they become fundamental rights. |
| **INTEREST OPTIMISATION** | The extent to which stakeholders are able to minimise the negative effects of a decision and maximise its positive effects in a way that takes into account the interests of the whole stakeholder group. |
| **EFFICIENCY OF COMMUNITY ENGAGEMENT** | The extent to which individuals are able to invest effort, success and the quality of relationships in running the activities that can be carried out at the collective level. |
| **PREVALENCE OF AI** | The extent of the living spaces, individual and social functions, the number of users in which AI performs essential tasks. |
| **QUALITY OF AI SERVICES** | The quality, breadth, diversity and error rate of the services provided by AI. |
| **DOMINANCE OF PROFIT LOGIC** | The extent to which human activities are motivated solely by profit motives. |
| **OTHER VARIABLES** | |
| **DIGITAL ADDICTION** | The extent of psychologically harmful dependence on digital solutions. |
| **PROTECTION OF THE DIGITAL SELF** | The extent to which systems are effective in ensuring that personal data created in the digital space cannot be misused. |
| **EXISTENTIAL AWARENESS** | The degree of awareness of human mortality. |
| **AUTONOMY OF HUMAN DECISIONS** | The ability of individuals to make informed, uncoerced decisions. |
| **QUALITY OF PUBLIC SERVICES** | The quality of all services that contribute to the achievement of community goals and well-being (e.g. health, education, culture, mobility). |
| **RESOURCE DEMAND OF AI** | The amount of energy, raw materials, other resources required to operate the AI alone. |
| **DECENTRALISATION OF THE USE OF AI** | The extent to which the use of systems incorporating artificial intelligence is dependent on conditions set by a single or a few providers. |
| **ACCESSIBILITY OF AI** | The extent to which a wide range of people in society can use AI. |
| **INCLUSIVITY OF AI** | The extent to which planning takes into account the specificities of different social groups. |
| **TRUST IN AI** | The extent to which people believe that AI will benefit them without putting them at risk, creating difficulties for them in the future, or misusing the data generated by its use. |
| **KNOWLEDGE OF AI** | The extent of individuals' knowledge of AI. |
| **WELFARE SURPLUS OF AI** | The extent to which the operation of the AI generates additional revenue, welfare surplus. |

| | |
|---|---|
| **TRANSPARENCY OF AI** | The extent to which the solutions, motivations, circumstances and conditions behind the resulting AI solutions are transparent to society. |
| **CUSTOMISATION OF AI** | The degree of personalisation of AI solutions and services. |
| **POLITICAL SUPPORT FOR AI** | The level of public support for AI, as expressed in policy interventions such as those that help AI to become more widely known and used |
| **DECENTRALISATION OF AI OWNERSHIP** | The extent of ownership of systems using artificial intelligence. |
| **JUSTICE OF THE REDISTRIBUTION OF THE AI CREATED SURPLUSES** | The extent to which the benefits of the creation, use and welfare gains of AI can be distributed by society to a wide range of stakeholders according to the ethical norms chosen |
| **HARMONY OF AI-HUMAN SYMBIOSIS** | The extent to which humanity is able to live with the AI around it in a way that is satisfying and does not conflict with individual/community aspirations and the environment provided by the AI. |
| **SELF-REALISATION** | The extent to which people can engage in activities that give them pleasure and make them feel fulfilled (not necessarily paid work in the current sense). |
| **SELF-DETERMINATION** | The extent to which individual rights are exercised consciously at the societal level. |
| **SELF-ORGANISATION** | The extent to which people are able to act together (loosely or more closely) as a cohesive group around a place of residence, work, learning or interests/interests, values, or a cause, so that the action is initiated and organised "on its own", by someone or someone-else, without external impulses. |
| **DECENTRALISATION OF NEED SATISFICERS** | The extent to which humanity is able to respond to the same human need from a variety of sources, by a variety of means, using different approaches. |
| **SOCIAL RESPONSIBILITY** | The extent of the commitment and obligation of individuals, members of society or communities to contribute voluntarily to actions in the collective interest. |
| **DEPENDENCE ON TECHNOLOGY** | The extent to which individuals, communities and society can meet their own needs without the availability of technology |
| **OVERCONSUMPTION** | At the individual level, the level of normal needs or average available resources, and at the collective level, the level of consumption above the Earth's carrying capacity. |


## Acknowledgements

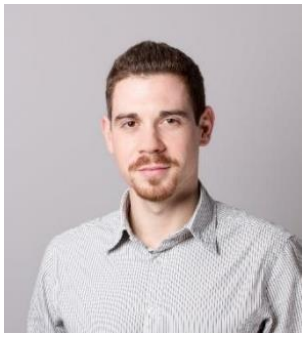
The authors would like to thank all the participants of the backcasting workshop for volunteering their time and efforts to this research and for sharing their very valuable ideas with us. We would also like to thank Róbert Pintér, who helped to recruit the participants and also facilitated groupwork in the workshops.

# About the authors

Alexandra Köves: ecological economist, graduated from Heriot-Watt University in Edinburgh and the Diplomatic Academy in Vienna. She spent more than 15 years working on social development in the field of EU development policy, a period which has led to her interest in alternative approaches to economics. She is currently Associate Professor at Corvinus University of Budapest. Her research explores the potential for an environmentally and socially sustainable future by moving beyond current paradigms. She is an expert in the backcasting method. Founder and editor-in-chief of the podcast series Green Equality and Economics for Rebels. E-mail: alexandra.koves@uni-corvinus.hu

Lilla Vicsek: Sociologist, graduated from the Budapest University of Economics and the Central European University (CEU). Since 2005, she has been conducting research on the links between technology and society. Her previous research projects have investigated certain social aspects of genetically modified crops, stem cell research and reproductive technologies, among others. Her current research focuses on visions of artificial intelligence. She is the leader of the OTKA research project "*Visions on artificial intelligence and society*". E-mail: lilla.vicsek@uni-corvinus.hu

Katalin Fehér: She received her PhD in Communication and Media Studies and habilitated at the Doctoral School of Philosophy of ELTE. She holds a Fulbright scholarship from the USA and is an Associate Professor at the University of Public Service. She is a post-doctoral MTA Bolyai fellow and a visiting professor at Masaryk University in the Czech Republic. Her international experience focuses on the United States, South East Asia and Scandinavia. Her broader research interests are in socio-cultural AI, and her narrower interests are in AI-driven and synthetic media. E-mail: feher.katalin@uni-nke.hu



Máté Fischer: Master's degree in Economics from Corvinus University of Budapest, Budapest, Hungary, currently PhD student at the University. He has been working on technology projects for 6 years , mainly in the field of digital marketing. He has been working at Vodafone Hungary Zrt. for 3.5 years as a business analyst and marketing technology expert. E-mail: mt.fischer@yahoo.hu

# References

P. Barbrook-Johnson, A.S. Penn, Systems Mapping. Palgrave Macmillan, Cham, (2022). https://doi.org/10.1007/978-3-031-01919-7

J. Bergold, S. Thomas, Participatory research methods: A methodological approach in motion. Hist. Soc. Res. (2012) 191-222.

J. Brown, D. Issacs, World café: Shaping our futures through conversations that matter. San Francisco: Berrett-Koehler, (2005).

J.C. Glenn, (2009). Futures Wheel. Futures Research Methodology (Version 3.0), AC/UNU Millennium Project, Washington, D.C.

J. Quist, P. Vergragt, Past and future of backcasting: the shift to stakeholder participation and a proposal for a methodological framework. Futures, 38(9), (2006) 1027-1045. doi:10.1016/j.futures. 2006.02.010

J. Robinson, Future subjunctive: backcasting as social learning. Futures, 35(8), (2003) 839-856.

J. Robinson, S. Burch, S. Talwar, M. O'Shea, M. Walsh, Envisioning sustainability: Recent progress in the use of participatory backcasting approaches for sustainability research. Technological Forecasting and Social Change, 78(5), (2011) 756-768.

M. Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf Doubleday Publishing Group, (2017).