Exploring the Potentials and Pitfalls of Artificial Intelligence-Driven Decision-Making

Zoltan Zakota
Partium Christian University, Romania

Abstract

As artificial intelligence (AI) becomes more deeply implied in everyday life, it takes a more prominent role in decision-making in every industry. As Joseph Fuller, a professor of management practice at Harvard Business School, said: "Virtually every big company now has multiple AI systems and counts the deployment of AI as integral to their strategy." Implicitly, decision-making capabilities are incorporated into their products; consequently, ethical concerns also gain importance. This paper presents some of the most critical issues of using AI in everyday decision-making. Starting from the three main concepts of AI, decision-making and ethics, it is a philosophical approach to the issues and biases raised by AI's overwhelming spread in everyday life.

Introduction

Today, artificial intelligence (AI) is everywhere in our environment. However, a superficial observer looking around our world sees almost the same landscape as a couple of decades ago: computer networks, the Internet, and all kinds of digital services. Deep down, radical changes have taken place during this time. AI has become ubiquitous, and its omnipresence seems to generate profound reactions at both emotional and rational levels, not only among experts but also among the general public.

Different levels of AI elements are integrated into more and more devices, and the number of those who also have decision-making skills is rapidly increasing. However, machine decision-making faces both manufacturers and regulators with questions that are very difficult to answer. As spectacular as the advances in scientific and technical fields are, as difficult it is for us to answer the ethical questions they generate.

In the following, the paper presents briefly the current state of this topic, starting from a theoretical basis and using the latest results in the field.

Defining AI

As Patrick H. Winston, one of the founders of AI, states in his classic book, "There are many ways to define the field of Artificial Intelligence." Hereafter, he continues with his definition: "Artificial intelligence is ... [t]he study of the computations that make it possible to perceive, reason, and act. From the perspective of this definition, artificial intelligence differs from most of psychology because of the greater emphasis on computation, and artificial intelligence differs from most computer science because of the emphasis on perception, reasoning, and action." (Winston,1993). His approach brings together the two main groups involved in developing AI: computer scientists and engineers interested in making computers more useful and psychologists, biologists, linguists, and philosophers interested in a better understanding of the foundations of intelligence.
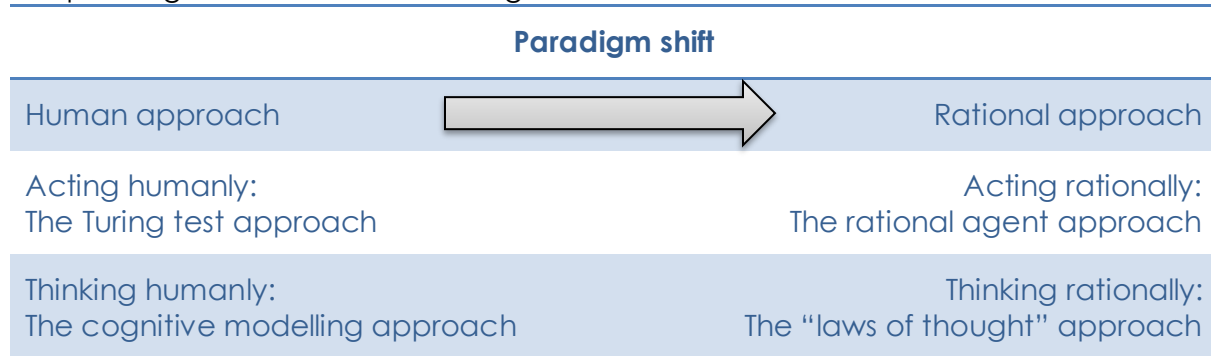
As for the renewed optimism of the turn of the millennium, it is well illustrated by the words of a Hungarian leader in the field: "What is usually called artificial intelligence is actually the thinking and scientific forerunner of computer science. If we did not live in a world full of messages, in which we have to find a good-sounding, even sensational name for everything, we could identify what we sell under the heading of artificial intelligence with computer science, primarily with its most advanced chapters. These chapters, if they work out in practice, will one by one be integrated into the general toolkit of computer technology and will become an unnoticeable part of commercial programs." (Futó, 1999). It is hard not to notice his position, which is opposite to that of Patrick H. Winston, by not separating computer science from AI but designating the latter as its cutting-edge field.

After three decades, one of the world leaders in the field designates AI as "systems or machines that mimic human intelligence to perform tasks and can iteratively improve themselves based on the information they collect." (Oracle, 2022). We can find a similarly anthropocentric definition on the Investopedia site: "[AI] refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind, such as learning and problem-solving." (Frankenfield, 2022).

Nowadays, this perspective is considered obsolete by many leading theorists and practitioners. Since the first "great" era of AI, we have witnessed a profound paradigm shift in the field from a human or rather anthropocentric approach to a rational one.

The next table summarises this shift based on (Russell & Norvig, 2021). Although this approach to AI is still widely used, we can now witness a return to the more anthropocentric perspective at a higher level.

Table 1
The paradigm shift in Artificial Intelligence

Paradigm shift	
Human approach	Rational approach
Acting humanly: The Turing test approach	Acting rationally: The rational agent approach
Thinking humanly: The cognitive modelling approach	Thinking rationally: The "laws of thought" approach

Source: Author's work

Nowadays, another much looser interpretation of the term is rapidly gaining ground, especially in the press and pop culture. According to this, the term *Artificial Intelligence* (AI) designates any software or hardware (usually a combination of them) capable of mimicking, to some degree, intelligent behaviour, e.g., chatbots, robots, and expert systems.

AI as the new digital environment

The above-cited words of Joseph Fuller (Pazzanese, 2020) express probably the most relevant characteristics of modern companies for which AI has become the new digital environment. AI has emerged as a transformative force, reshaping the digital landscape and revolutionising the way we interact with technology, influencing various aspects of our daily lives and presenting both opportunities and challenges. The rise of AI has ushered in a new era of digital transformation, with technologies being no longer limited to experimental projects but have become integral to numerous applications, from personalised recommendations to autonomous systems. "The global AI adoption rate grew steadily and now is 35%, a four-point increase from the year before. And in some industries and countries, the use of AI is practically ubiquitous." (IBM, 2022, p. 2).

AI is evolving into the new digital environment, not only by transcending traditional computing paradigms and driving significant changes across multiple domains but also by dismantling geographical barriers and deepening globalisation. In a 2018 discussion paper on the impact of AI on the world economy, the McKinsey Global Institute identified four major trends at macro and global level (Bughin et al., 2018):

- AI has a large potential to contribute to global economic activity.
- The economic impact may emerge gradually and be visible only over time.
- A key challenge is that the adoption of AI could widen gaps between countries, companies, and workers.
- How companies and countries choose to embrace AI will likely impact outcomes.

Five years later, in their Technology Trends Outlook 2023, they identified the AI revolution as the major trend of the near future. With the help of applied AI, with its "capabilities, such as machine learning (ML), computer vision, and natural language processing (NLP), companies in all industries can use data and derive insights to

automate processes, add or augment capabilities, and make better decisions." They estimate the potential economic value of applied AI to be somewhere between 17-26 trillion USD and the proportion of responding organisations adopting AI rising from 20% in 2017 to 50% in 2022 (Chui et al., 2023).

The integration of AI into industries has resulted in increased efficiency, productivity, and innovation. Fields like healthcare, finance, manufacturing, and transportation are adopting AI-powered solutions to optimise processes, analyse data, and make better decisions, while AI-driven automation is streamlining operations, freeing up human resources for more complex tasks, and enhancing overall performance.

AI has enabled hyper-personalisation in digital services and products. Recommendation systems, chatbots, and virtual assistants use machine learning algorithms to understand individual preferences and provide tailored experiences. As AI continues to gather data and learn from user interactions, the digital environment becomes increasingly intuitive and user-centric. One area where this huge potential of AI is just beginning to emerge is personalised education and training.

AI's integration into the digital environment has also extended to creative domains. AI-generated art, music, and literature are gaining popularity, challenging conventional notions of creativity. While AI-generated content can inspire and assist human creators, it also raises questions about originality, authorship, and the nature of creativity.

As so many times in human history, the fast development of AI we have witnessed in the last couple of decades was induced by two main factors: scientific and technological development, on the one hand, and economic opportunity, on the other. As presented in (Oracle, 2022), the driving factors of the rapid adoption of AI are:

- affordable, high-performance computing capability is readily available;
- large volumes of data are available for training;
- applied AI delivers a competitive advantage.

AI-driven decision-making

The concept of decision-making is inseparable from that of information, and classic definitions also refer to this. As Jay Wright Forrester states in his fundamental work: "Management is the process of converting information into action. The conversion process we call decision making." (Forrester, 1965, p. 93) This definition reflects primarily a management's point of view, but we can take into consideration an even more general interpretation. As Attila Chikán, a former professor of corporate economics in Budapest, pointed it out, a decision is a "choice between action options." (Chikán, 2008). Generally speaking, decision-making is a cognitive process resulting in the selection of a belief or a course of action among several possible alternative options. This process could be either rational or irrational. In other words, it is a reasoning process based on assumptions of values, preferences and beliefs of the decision-maker.

Although the previous definitions may be satisfactory at the level of everyday life, for example, within an organisation, if we want to deepen and/or nuance it, it is more useful to use a psychological approach. "In psychology, decision-making […] is regarded as the cognitive process resulting in the selection of a belief or a course of action among several possible alternative options. Decision-making is the process of identifying and choosing alternatives based on the values, preferences and beliefs of the decision-maker. Every decision-making process produces a final choice, which may or may not prompt action. Decision-making can be regarded as a problem-solving activity yielding a solution deemed to be optimal or at least satisfactory. It is,

therefore, a process that can be more or less rational or irrational and based on explicit or tacit knowledge and beliefs. Tacit knowledge is often used to fill the gaps in complex decision-making processes. Usually, both of these types of knowledge, tacit and explicit, are used together in the decision-making process." (IJSCP, 2023).

The ultimate goal of decision-making is problem-solving, i.e. "a creative process to resolve an inappropriate situation." (Chikán, 2008). People are constantly facing decision-making situations in all areas of their lives. Decisions are usually based on the data available to the decision-maker at the time of decision-making. The more and better the data, the better decisions can be made, but with all this data comes a problem, namely, how can human brains possibly process it all? Moreover, this is where AI comes into play.

AI is becoming an essential component of decision-making processes, particularly in areas like finance, law, and healthcare. While AI can provide valuable insights and optimise decisions, it should be seen as an aid rather than a replacement for human judgment. The human-AI partnership must be carefully managed to ensure accountability and prevent undue reliance on automated systems.

As AI-powered systems become more complex, vulnerabilities and attack surfaces increase, so AI's growing influence demands robust security measures to protect against potential threats. Building trust in AI technologies among users is essential for their widespread acceptance, necessitating transparent practices and explainable AI models.

AI-ethics

According to a maybe simplistic but definitely practical interpretation, ethics is a set of moral principles which help people discern between right and wrong. Based on this, we can define AI-ethics as a set of guidelines that advise on the design and outcomes of AI. At least, this is the standpoint of many developers and other specialists in the field when dealing with societal concerns raised by AI issues like job displacement, privacy breaches, bias in decision-making algorithms, or the concentration of power in tech giants. Ethical considerations are crucial in ensuring that AI technologies are developed and deployed responsibly, preserving human rights and avoiding harm. Nowadays, tech leaders try to set out lists of primary concerns regarding AI, like the following one (IBM Cloud Education, 2021):

- Technological singularity,
- Impact on jobs,
- Privacy,
- Bias and discrimination,
- Accountability.

In the short term, the fear of the expected transformation of the labour market is probably the determining element of the attitude of the general public towards AI. Expectations cover the entire emotional spectrum, from deeply pessimistic, anti-capitalist viewpoints to optimistic utopias envisioning a wonderful future. A good example of the former is Jason Resnikoff, who states: "Globalised production, the growing power of Silicon Valley, innovations in robotics and artificial intelligence, the ever-growing economic and political importance of the Internet, and the weakness of unions and nation-states alike to check neoliberal capital once again made 'automation' an urgent concern. […] In this environment, many ask, 'What will the machines do to us?' rather than, 'What do we want our machines to do?'" (Resnikoff, 2021).

On the other hand, Daniel Susskind convincingly argues that the situation is indeed different now, as AI will put most people out of work in the coming decades. (Susskind, 2020) Between the two, this year's OECD report probably represents the most realistic point of view: "Despite the renewed worries about a jobless future, the impact of AI on job levels has been limited so far. We are at a very early stage of AI adoption that is generally concentrated in the largest firms that are often still experimenting with these new technologies." (OECD, 2023, p. 5).

The other great fear of the masses is the technological singularity, a point of no return in the future when technological advancement becomes uncontrollable and irreversible, inducing unpredictable and unwanted changes in human civilisation. According to Stanislaw Ulam, the concept of singularity was first used in the technological context by John von Neumann when referring to „the accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue" (Ulam, 1958, p. 5). The term gained popularity in the essay of Vernor Vinge (Vinge, 1993) and later in the book of futurist Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, predicting singularity by 2045 (Kurzweil, 2005). We should also mention Murray Shanahan's work here (Shanahan, 2015).

To ensure that AI continues to enhance the digital environment positively, responsible AI development is imperative. Governments, industry leaders, and researchers must collaborate to establish ethical guidelines, regulations, and best practices. Transparency, fairness, and accountability should be prioritised to build trust and prevent the misuse of AI technologies. Leaders in the field compete in elaborating principles for AI ethics, as it does, e.g., IBM's set of *Principles for Establishing AI ethics* (IBM Cloud Education, 2021):

- Respect for persons: recognising the autonomy of individuals and upholding an expectation for researchers to protect individuals with diminished autonomy.
- Justice: fairness and equality; who should benefit from experimentation and machine learning?
  - Equal share
  - Individual need
  - Individual effort
  - Societal contribution
  - Merit
- Beneficence: "do no harm" - counterbalancing the amplification of biases around race, gender, and political leanings by algorithms, despite the intention to do good and improve a given system.
- Governance: companies can leverage their existing organisational structure to help manage ethical AI (e.g. data collection, data standardisation).
- Explainability: eliminating the „black box" ambiguity around the model assembly and model outputs by generating a "human understandable explanation that expresses the rationale of the machine".

Conclusion

The rapid pace of AI development promises an exciting future for the digital environment. Advancements in natural language processing, computer vision and reinforcement learning will open new possibilities for AI applications. As AI becomes more ubiquitous, interdisciplinary collaboration will be crucial to tackle challenges and harness its full potential.

AI has emerged as the new digital environment, revolutionising industries, transforming user experiences, and shaping society's future. However, this transformation also comes with challenges, necessitating a balanced approach to AI development and deployment. By prioritising responsible AI practices and fostering collaboration, we can create a digital environment where AI enhances human lives while respecting ethical principles and values.

As we can see from the previous arguments, the very fast and very steep rise of AI in recent years brings with it many advantages, but at the same time, it also generates serious dangers. Humanity is facing an unprecedented situation, the outcome of which is still completely uncertain, and our experience so far has not helped much to solve it. What we can hope for is that we can successfully recover from this by placing ourselves on a strong ethical foundation.

References

1. Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). Notes from the AI Frontier: Modeling the Impact of AI on the World Economy. McKinsey & Company.
2. Chikán, A. (2008). Vállalatgazdaságtan /Business Economy/ (4th revised & expanded ed.). AULA.
3. Chui, M., Issler, M., Roger, R., & Yee, L. (2023, 07 20). Technology Trends Outlook 2023. Retrieved 08 04, 2023, from McKinsey Digital: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/mckinsey%20technology%20trends%20outlook%202023/mckinsey-technology-trends-outlook-2023-v5.pdf
4. Forrester, J. W. (1965). Industrial Dynamics. Cambridge, Massachusetts: The M.I.T. Press.
5. Frankenfield, J. (2022, 06 06). Artificial Intelligence: What It Is and How It Is Used. https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp
6. Futó, I. (1999). Előszó /Foreword/. In I. Futó (Ed.), Mesterséges intelligencia /Artificial Intelligence/. Budapest: Aula.
7. IBM. (2022, May). IBM Global AI Adoption Index 2022. https://www.ibm.com/downloads/cas/GVAGA3JP
8. IBM Cloud Education. (2021, 03 18). AI Ethics. IBM: https://www.ibm.com/cloud/learn/ai-ethics
9. IJSCP. (2023). Decision-making. (Longdom Publishing) Retrieved 02 18, 2023, from International Journal of School and Cognitive Psychology: https://www.longdom.org/peer-reviewed-journals/decisionmaking-27214.html
10. Kurzweil, R. (2005). The singularity is near: when humans trascend biology. Viking.
11. OECD. (2023). OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market. (A. Bassanini, & S. Broecke, Eds.) https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2023_08785bba-en
12. Oracle. (2022). What is AI? Learn about Artificial Intelligence. https://www.oracle.com/artificial-intelligence/what-is-ai/
13. Pazzanese, C. (2020, 10 26). Great promise but potential for peril. https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/
14. Resnikoff, J. (2021). Labor's end: how the promise of automation degraded work. Urbana, Chicago and Springfield: University of Illinois Press.
15. Russell, S. J., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (Fourth ed.). Hoboken, NJ: Pearson Education, Inc.
16. Shanahan, M. (2015). The technological singularity. Massachusetts Institute of Technology.
17. Susskind, D. (2020). A World Without Work - Technology, Automation, and How We Should Respond. New York: Henry Holt and Co.
18. Ulam, S. (1958, May). John von Neumann 1903-1957. Bulletin of the American Mathematical Society, 64(3)
19. Winston, P. H. (1993). Artificial Intelligence (Third ed.). Addison-Wesley

About the author

Zoltan Zakota is an electrical and environmental engineer, as well as a lecturer at the Partium Christian University in Oradea, Romania. He currently teaches computer science and economics subjects. In addition, he teaches computer science and electrical engineering subjects at the Faculty of Engineering of the University of Debrecen, Hungary. Over the years, in addition to education, he also worked in the private and civil spheres. He participated in many domestic and international projects, mainly in the field of education and research. His main areas of interest are the information and knowledge-based society and the impact of ICT on society, economy and education. The author can be contacted by email at **zzakota@gmail.com**.