ECONOMY AND FINANCE • VOL. 12. ISS. 1 • MARCH 2025 • DOI: 10.33908/EF.2025.1.5

OPERATIONAL RISKS CAUSED BY AI USE AT WORK AND THEIR MANAGEMENT IN PROFESSIONAL LITERATURE

Péter Juhász[1]

ABSTRACT

The increasing use of artificial intelligence (AI) at workplaces carries major risks, appearing at both technical, legal, and structural levels. Risks include – among others – GDPR concerns and distortions of decision-making while they also cover blurred responsibilities and changes in human jobs. Based on a comprehensive review of professional literature, this study defines categories of AI-specific operational risks with particular attention to ethical dilemmas, regulatory challenges, and the impact on organisational effectiveness. The analysis sheds light on the paradoxes generated by AI that complicate corporate decision-making and risk management. Finally, the study proposes proactively managing AI-related risks; the most important ones include improved transparency, application of adaptive risk management models and ongoing improvement of the regulatory framework. The findings show the organisational integration of AI is not simply a technological but also a strategic and structural issue that requires a long-term approach.

JEL codes: G28, G32, M15, M51, O33, O34

Keywords: artificial intelligence, risk management, corporate operation, adaptation, human resources

1 *Péter Juhász,* PhD, CFA, Associate professor, Budapest Corvinus University, Institute of Finance, E-mail: peter.juhasz@uni-corvinus.hu.

1 INTRODUCTION

Although the history of AI goes back several decades, an explosive growth of AI-based applications commenced at the beginning of the 2010s when deep learning exceeded earlier machine learning in terms of efficiency (Krizhevsky–Sutskever–Hinton, 2012). The new technology has boosted natural language processing (NLP) while computers' computational performance has increased radically. This has become most apparent in the development of graphic processing units (GPUs) when NVIDIA of the USA started to focus on the field. AI-specific machines, for instance, Tensor Processing Units (TPUs) by Google, further advanced improvements, as large neural networks could be generated effectively.

Development, however, could not have taken place had it not been for the huge body of data generated by social media and business applications over earlier decades which could be used for training. Meanwhile companies gained access to immense computational capacities by means of cloud service providers (AWS, Google Cloud, Microsoft Azure) without having to build their own infrastructure.

It was probably Google that put an avalanche in motion when it presented its model BERT (Bidirectional Encoder Representations from Transformers) in 2018, which significantly improved the accuracy of translation and text generation tasks (Devlin et al., 2018). The real breakthrough, however, came with the debut of OpenAI GPT-3 in 2020. As text generation continued to improve, a wide range of applicable models appeared including AI-based chatbots, virtual assistants and content developers.

Giants entered the market: Google (DeepMind), Meta (Llama) running Facebook, Anthropic (Claude) and Perplexity AI partially financed by Nvidia and Jeff Bezos but, in theory, independent. The entry of the Chinese DeepSeek has shaken up the market recently (CBSNews, 2025). Compared to its competitors' 16-thousand GPU capacity requirement, the model R1 is alleged to need a mere 2000 GPU to achieve remarkably similar performance. The sudden popularity of R1 did not only reduce the value of Nvidia shares by 17 percent, but it was also a clear sign the market is willing to accept any novelty, as no strong brand loyalty has been formed yet. While several countries launched investigations to assess the security risks of R1, the operators of AI-based applications have started to switch to the Chinese solution since its price is only 0.6 percent of that charged by OpenAI (Reuters, 2025).

In line with the development of AI, IT companies have started to build it into their products, which has resulted in major changes in industries such as finance, healthcare, and the motor industry (Bughin et al., 2018). AI-based solutions have become an attractive target for venture capital. By now, the technology has even been made available for free thanks to some free versions and to

the AI assistants built into the new versions of software products traditionally sold in high volumes. However, most active AI users are unaware of the significant differences between traditional and AI-based solutions. (*Table 1* displays the most important ones).

All that, on the other hand, means that AI and its daily use affect the operations of a growing number of companies, while they are not necessarily prepared for the new risks associated with the new technology. This study has been built on a comprehensive review of the professional literature to present the challenges you will have to face and how to manage the new risks effectively.

First, let us see what companies hope for when they introduce AI-based solutions; next, there is an analysis of the applicability of classical risk management models. After that, the operational risks of using AI and their categories are presented, followed by a review of options for risk management. However, organisations must solve a number of AI-related paradoxes so that AIs can be successfully introduced and new methods of risk management implemented. Finally, the study lists all of the above, summarises its findings and offers some proposals.

2 IMPACT OF AI ON CORPORATE OPERATIONS

AI-based applications can have an impact on several areas of corporate operations simultaneously. For instance, improvements in the efficiency of data analytics tools can reshape the worlds of controlling, sales and logistics, healthcare diagnostics or stock exchange trading (Bughin et al., 2018). AI-based systems cannot only be used to reduce operating costs, but decision-making can also improve, or new vistas can open up for innovation (Davenport –Ronanki, 2018). Such tools are, in fact, available for all undertakings as their costs are declining. Thus, you can expect the number of enterprises affected by fundamental changes to grow.

Table 1
Comparison of traditional and AI-based systems

	Traditional software tools	AI-based solutions
Areas of use	Structured, rule-based environments (banking systems, ERP, databases)	Complex, unstructured environments (image recognition, autonomous systems)
Reliability	Highly dependable, deterministic, and rule-based	Less dependable, works on probability, subject to data quality
Decision-making	Based on pre-determined logic and strict rules	Uses statistical models and probability conclusions
Flexibility	Rigid, requires reprogramming for changes	Adaptive, can learn from data, develops continuously
Transparency	Clear logic, easy to audit	Models are often not transparent (black box), difficult to interpret
Error management	Errors are unambiguous, easy to correct by debugging	Errors are hidden, subject to data, and difficult to trace back
Scalability	Scaled subject to hardware and architecture	Scaled subject to data availability and computing capacity
User interaction	Runs on structured inputs (menus, forms, buttons)	Can process natural language, images, and complex inputs
Adaptive capability	Requires manual updates and re-programming	Can develop on its own through learning and finetuning
Performance measurement	Based on accuracy and pre-defined test cases	Only environment-specific robustness is measurable
Security risks	Hacker attacks	Distortions and model poisoning
Ethical considerations	Few ethical worries as pre-defined rules are followed	Ethical issues arise (bias, equity, accountability)
Data dependency	Runs on structured pre-defined data formats	Learns from high volumes of (often) unstructured data
Regulatory compliance	Easy to regulate as it is deterministic	Difficult to regulate as its decision-making changes all the time and may involve distortions

Source: own design

AI can help develop marketing content, business decision-making or customer service. In one case, actual complaint management time using Claude AI was shortened by 87 percent (Reuters, 2025). Data analysis capabilities using AI can improve human resources management too (Financial Times, 2024). It cannot only boost developing training materials but also explore internal organisational skills and knowledge; it can pre-select candidates applying for a position or speed

up the selection of leaders more successfully. In terms of logistics, AI-based solutions can be trailblazers in optimising routes, planning and recording inventory or predicting demand (Ringby. 2025).

In healthcare, AI has revolutionised patient care, particularly imaging diagnostics. An international study has shown the accuracy of deep learning algorithms exceeds the performance of radiologists in breast cancer screening. In the finance sector, AI-controlled tools provide outstanding performance in automated risk analysis, detection of frauds and algorithmic trading. According to a study (Fuster et al., 2022), financial institutions applying AI technology make faster and more accurate credit scoring decisions reducing in that way non-performance risk and improving stability.

To sum up, it seems AI-based solutions are highly disruptive: they bring about radical changes, rendering a lot of earlier solutions and technologies obsolete. (*Table 1* is a comparison of AI-based and traditional solutions). What is more, in contrast to similar huge leaps in earlier ages (steam engine, automotive, computer, internet) users of AI do not apparently have to learn new capabilities: AI can adapt to the user, those inexperienced in text generation, imaging or resource evaluation are also able to reach at least medium-level results quite quickly. It can be a great help for companies where no skills are available in certain fields or the time to be spent on a task is too short.

As AI technology improves, the difference between products manufactured by a low-skilled workforce or by a machine will be difficult to point out. In some fields, the quality advantage of work produced by medium skilled workforce seems to be disappearing in some areas. AI can threaten traditional jobs as automation spreads, which can cause workplace polarisation and increase economic inequality if people of lower abilities find themselves at a disadvantage (Hassel–Özkiziltan, 2023). AI has a rather selective impact on the world of labour: it can quicken up and simplify work for highly skilled people, while it can even render medium-skilled labour superfluous, while low-skilled or no-skill labourers may find themselves in a situation where their immediate managers have been replaced by AI.

Watching such trends many business owners and top managers may rightfully think the time has come to open up space for AI-based solutions in the operation of their companies. Still, rather than haste, you should prepare well for implementation. It is already clear the use of AI may come hand in hand with the appearance of major and novel operational risks.

3 CONCEPTUAL FRAMEWORK / TERMINOLOGY

As a first step to categorise operational risks caused by AI, one should review the traditional frameworks of risk management. The most frequently used ones include the Risk Management Framework (RMF), the Three-Lines Defence (TLD) model, COSO corporate risk management (ERM) framework (COSO, 2017) and the Swiss cheese model (SCM). See an excellent summary of IT models of risk and security by Csáki (2023).

RMF focuses on uninterrupted risk assessment and monitoring, while TLD emphasises responsibilities: operations management (first line), risk management and compliance functions (second line), while internal audit (third line) provide comprehensive risk control. COSO ERM has integrated risk management and strategic planning, emphasising the importance of risk awareness in decision-making processes. On the other hand, the Swiss cheese model has moved its focus onto human interactions contributing to the development of risks.

Traditional models, however, have limitations in terms of applicability to AI-related risks. Former risk rating rules must be replaced with AI-specific taxonomies (MIT, 2024); static risk assessment models must be substituted with dynamic ones requiring uninterrupted monitoring and adaptive management solutions. The context-mechanism-risk (CMR) model already in use must be supplemented with the application environment, operating processes, and risks of AI.

Cummings (2024) proposes using TAIHA – a version he has developed and adjusted to AI activity – instead of the traditional SCM. In it, the focus is on human activity related to establishing and regulating AI-based solutions, however, it fails to cover the user-AI interactions that are central to other models. To settle the issue, the National Institute of Standards and Technology under the US Department of Commerce published a risk management framework explicitly adjusted to AI in the summer of 2024 (NIST, 2024) that has identified twelve AI-specific risk types.

4 AI-SPECIFIC OPERATIONAL RISKS

Traditionally, risks are categorised into at least three dimensions: cause, form of appearance and origin. For instance, a sudden rise in raw material prices (cause) can result in liquidity problems (appearance), and you are exposed to the risk because you operate in the given industry (origin). Categorisation is key because it will be the clue for risk management. If you only focus on the type of the risk, price fluctuation as market risk would be mitigated by applying derivative financial products, while the resulting liquidity issue would be remedied by holding

excess funds, finally industrial risk causing exposure might be managed by diversifying your activities, while – in fact – they are different dimensions of the same phenomenon.

Many articles and studies have been published recently on AI-related operational risks, which have recommended different ways of categorisation. For instance, you can see the causal taxonomy by MIT (2024) in *Table 2*, while Csáki (2024) has proposed a taxonomy in *Table 3* to categorise narrow AI risks. (Narrow AI means data-driven model-based intelligent systems typically built on machine learning methodologies.)

Table 2
Causal taxonomy of AI-related risks

Category	Level
Decision maker	Human AI Other
Intention	Intentional Unintentional Other
Time of occurrence	Prior to installation of AI Following installation of AI Other

Source: MIT (2024, p. 5.)

Table 3
Dimensional review of narrow AI-related risks

Place where risk appears		Error	Weakness
Technology	Data	Distorted (bias) Non-pure data Non-representative training data Limited or too large dataset Erroneous data transfer from earlier model Lack of data update	Misleading self-image in data Lack of alternative options Inappropriate management of size, speed, or complexity
	Model	Inappropriate utility function Inappropriate mapping of expert knowledge Erroneous model transfer Lack of model update Unsuitable proxies applied	Lack of explainability Lack of transparency Stakeholders' ethical preferences misunderstood
Organisation		Poorly targeted or asked questions Unsuitable context of use Faulty assumptions (what it means for the organisation)	Arrogant algorithm Average user cannot understand

Impact	Error	Weakness
Relationship between individual and organisation	Control lost over system Application incompliant with regulation	Unplanned side effect Autonomy related (responsibility) risk Accountability lost Behaviour affected
Relationship between individual and society	Too fast technological innovation Individuals lose their special position and do not feel to be useful (AI is better) False assumptions (what others think of him/her) AI applied for dangerous or harmful purposes Unleashed autonomous weapons Faulty regulation	Social isolation Moral relativism "Playing God" Loss of jobs Inequal incomes Growing economic differences Social tension because jobs are transformed

Source: based on Csáki (2023, p. 44.)

Using another approach risks can be analysed in three main dimensions: (I) technological, (II) organisational, and (III) social-ethical risks. Technological risks include issues of data protection, lack of transparency of systems and faulty models. Organisational risks cover warped decision-making processes, human labour substituted and legal compliance. Finally, social, and ethical risks include AI-related discrimination, spreading misleading information and psychological effects at workplaces. According to Hassel and Özkiziltan (2023), one should mainly differentiate AI-related risks by whether they exert a direct or indirect impact on work.

MIT (2024), on the other hand, proposed categorisation by forms of appearance. Next, here is a review of the risks identified in the professional literature particularly the twelve-part topology by NIST (2024).

A. Discrimination and toxicity

A.1 Unfair discrimination and misleading presentation. AI-based solutions are prone to have different prejudices (primarily due to the hidden features of the documents used for their training). A good example is the case of Amazon. Its AI-based selection system ranked women applying for IT jobs lower because – based on data from earlier years – the company used to prefer employing men to women in such positions (Dastin, 2018). Buolamwini and Gebru (2018) pointed out the accuracy of face recognition systems can be quite varied for different racial groups, which may result in discriminative situations. NIST (2024) mentions under a separate heading that intentionally toxic and discriminative content can easily be generated using AI-based solutions.

A.2 Sharing harmful content. It can occur that AI unexpectedly exposes users to toxic content: one has come across hate speech, encouraging suicide, support of illegal actions or (child) pornography. In addition to the above, the taxonomy of NIST (NIST, 2024) mentions a separate risk category, i.e., that knowledge related to chemical, biological, radiological, and nuclear weapons (VBSN-CBRN) and other dangerous materials are made available for large groups of people.

A.3 Inequal group performance. AI is prone to generating different results or decisions based on users' belonging to different groups mainly because of faulty system design or distorted materials used for training. In addition, AI may offer people having extremist views contents reiterating those views, which may further strengthen prejudices already existing.

B. Data protection and security

B.1 Violates data protection by obtaining, leaking, or correctly figuring out sensitive information. AI can learn sensitive information and share it with others without the agreement of those involved. Deepfake videos are an extreme example of it. A classic example is the case of Clearview.AI: the company developing face recognition technology has collected over three billion photos mostly from social media, while even the FBI's face recognition database contains 411 million photos only. A number of legal proceedings were launched against the company for illegally collecting personal data, although six hundred law enforcement organisations also used the solution. Finally, a court decision made in 2024 ordered payment of USD 52 million to the injured parties, which resulted in the effective bankruptcy of the company, with an estimated value of USD 225 million at the time. To escape insolvency, the company promised to pay the injured parties with shares when it was listed on the SE later (Hill, 2021 and 2024).

The NIST taxonomy (2024) categorises damage caused to intellectual property under a separate heading. Different content materials protected with copyright and related rights may get included in content generated by AI with no licence or marking. It can undermine the lawful operation of society and reduces people's motivation to generate such creative content, which hinders human development.

Hassel and Özkiziltan (2023) describe collecting and monitoring the personal data of a larger-than-ever group of employees. Companies obligating their employees to wear RFID badges provided with different sensors can collect data about their employees' movements, chatting habits or social relations. What is more, replacing certain HR functions with AI-based applications can lead to situations when people's lives can be ruined because of some insufficiently objective system.

So termed "algorithm-based management" can annihilate the border between private life and the workplace and can violate civil rights. Continuing the train of thought raised by the study, it can happen in the world of personalised workplace punishment or reward that employees will not want to work well but in the manner expected by AI. However, working ideally as per AI does not necessarily match the interests of the shareholders or of the company; also, it is not sure everybody can work best if identical patterns must be followed. The above recalls memories of the age of scientific management using Taylor's gesture analysis (Krisztián–Nemeskéri, 2014.).

B.2 Vulnerability of AI systems and security attacks. As any other IT system, AI can be vulnerable to IT attacks. One can envisage unwanted influencing of the system or leakage of the data stored.

Although it is not mentioned separately in the taxonomy, Domokos and Sajtos (2024) among others have pointed out that AI-based systems are operated by just

a few large market players globally because of the huge demand of resources involved, so the organisational integration of AIs can increase third party risk, and the vulnerability of systems cannot be managed internally any longer. According to a survey by the Boston Consulting Group, 55 percent of AI errors appear with tools produced by third parties (Cogent Infotech, 2024). This can be a particular challenge in the financial service provider industry.

C Disinformation or misleading information

C.1 Sharing false or misleading information. AI can generate and share false or misleading information, which may cause harm to users. In addition, NIST (2024) topology separately categorises (I) confabulation or hallucination, when AI states erroneous or false facts it has generated as real in a convincing, definite manner, and (II) when it reiterates different stereotypes, prejudices and system-wide distortions mentioned above.

C.2 Pollution of the ecosystem of information and distortion of reality perception. The appearance of content materials on the worldwide web generated by AI on false facts causes pollution while providing content in line with users' preconceptions can generate a bubble around those users who, in turn, will not be able to correct their views. This is connected to the integrity of information given a separate heading in the NIST (2024) system. It means that distorted answers adjusted to users' preconceptions reach the worldwide web and can later be confused with facts and human opinions, which will further increase uncertainty and distortions in the learning databases of new AI systems. A good example of such risk is a deepfake photo from 2022 "depicting" Volodymyr Oleksandrovich Zelensky, in which the Ukrainian president seems to capitulate in the war against Russia (Pearson–Zinets, 2022).

D Malignant actors and abuses

Szabó (2023) has pointed out that, according to data published by the Interpol and Europol, AI has already changed the nature of crime. Committing criminal actions and access to formulas of highly dangerous materials causing much damage as well as misleading the authorities have become easier, in addition, the evidentiary procedure in criminal cases has also become more cumbersome because of deepfake technology.

D.1 Disinformation, mass surveillance and influencing AI systems controlled by a third party can be used for demagogy, organisation of misleading campaigns or surveillance of the users.

D.2 Cyber-attacks, arms development or use, mass damage. AI-based systems can be used for cyber-attacks, or the development of the tools needed for them.

D.3 Fraud, rip-offs, and targeted manipulation. AI systems can help different types of criminals and can increase the number of unintentional crimes, such as plagiarism.

E. Human-computer interaction

E.1 Users' excessive reliance on AI or dangerous use of AI. Excessive material or emotional reliance on AI may cause damage or offer points of attack to others. Maybe the best-known example here is the proceedings launched against Tesla by the US National Highway Traffic Safety Administration. The proceedings were closed in 2017 with no measures. However, investigations were restarted in 2021 following several accidents caused by co-pilot cars, which has led to recalling two million cars and software updates. NHTSA launched another investigation in October 2024 because of further accidents (Shepardson–Jin, 2021, Waltz, 2024). A separate heading in the NIST taxonomy (2024) is devoted to risks due to attributing human features to AI, such as psychological problems.

E.2 Decline of human independence and decision-making capability. Substituting man-made decisions with AI may lead to excessive reliance of the organisations on AI while human traits and emotions can disappear from the decisions. AI-controlled HR decisions may be inhuman, the management responsibility of human activities organised, planned, and supervised by AI is unclear.

The technostress impact (Ragu–Nathan, 2008) missing from the MIT taxonomy also belongs there. For instance, Lestari et al., (2023) have proved in their survey of fast-food restaurants that AI-related skills of employees have increased the level of technostress. Anxiety because of the appearance of new technologies has had an adverse effect on the performance of service staff. In other words, simply the existence of AI and its appearance at the workplace had an impact on human work performance. It is called indirect impact in the typology by Hassel and Özkiziltan.

F. Socio-economic and environmental damage

F.1 Power concentration and unfair distribution of benefits. Too much global power can be concentrated with groups controlling AI.

F.2 Increasing inequalities and decline of employment quality. The spread of AI-based applications may significantly reduce certain groups' chances to find employment, while the operators of AI will gain financially. According to a survey by CFA Survey in the US, 58 percent of the companies asked expected the corporate application of AI to result in quality improvement, 49 percent in increased

output, 47 percent in reduced labour costs and 33 percent in total substitution of their employees (Egan, 2024).

F.3 Economic and cultural degradation of human efforts. The part played by AI in creative activities (copywriting, programming, graphic arts) may have an adverse effect on human creativity, it can depreciate human efforts and lead to the establishment of a globally homogeneous culture.

F.4 Harmful dynamics in competition. The fast development and fierce competition of AIs may encourage/drive developers to market solutions that are faulty or have not been properly tested.

F.5 Government failures and regulatory deficiencies. (Many topologies take this as a separate main class, not least because they can be managed using different tools and actors.) Inadequate regulations may pave the way for AI-related misuse and hamper risk management. It should be noted this class fails to meet the building requirements of the original taxonomy. In this case, it is not AI that causes risk, but inadequate regulations impact both developers and AI. The other headings of the list cover the risks appearing as a result.

F.6 Environmental damage. The huge carbon print of AI systems and their energy consumption contribute to the decay of the natural environment. On the other hand, the appearance of DeepSeek is a promising phenomenon, as its computational needs are much lower to achieve performance similar to its competitors.

G Security, errors, and limitations of AI systems

G.1 AI can strive to reach its own goals that can be contrary to human values and goals. One can imagine that AI will manipulate users based on erroneous conclusions contrary to the interests of humanity.

G.2 Appearance or generation of dangerous capabilities of AI. Provided AI systems can directly influence the physical environment, the damage caused by faulty AI or its malignant manipulation can increase by an order of magnitude.

G.3 Deficiencies of capabilities and unreliability. If AI systems become unreliable under certain conditions, they can cause serious damage in critical systems. With respect to an accident caused by an autopilot Uber vehicle, the US National Transportation Safety Board identified the cause as insufficient tests and the lack of proper safety devices (NTSB, 2019). Similarly, faulty high-frequency trading (HFT) algorithms are usually blamed for "flash-crash" occurrences recently found on financial markets, when the price of a product goes into free fall in a few seconds' or minutes' time and then climbs back to its original level (Majumder–Yashraj, 2024). Robots collaborating with people (co-bots) may cause accidents

and injuries to people, while faults in AI-based medical instruments may lead to false diagnostics and treatment.

G.4 Lack of transparency or interpretability. If AI decisions lack transparency, trust in the decisions weakens, correcting errors and the accountability of those responsible may suffer. This may be particularly severe related to the control of critical systems, in healthcare and financial applications (Domokos–Sajtos, 2024).

G.5 The well-being and rights of AI. As the capabilities of AI develop, the issue of the systems becoming legal entities gets into the limelight. The minimum of their well-being might have to be identified, which will open up new ethical issues and lead to the appearance of new kinds of risks. It has already been proved by science (Yin et al., 2024) that AI responses to requests (prompts) worded nicely are of higher quality, but the optimum level of nicety is different from one language to the next. You do not know if an AI is more prone to hallucinate or give wrong answers in response to a rude or impolite question, or whether an AI-based system is able to have positive feelings towards one user but negative ones towards others.

It should be noted the MIT taxonomy (2024) has its focus on the form of appearance and source of risk, while many traditional sources focus on the nature of damage. For instance, all of the formers may cause reputational risk (Holweg et al., 2022) if a company's assessment is damaged due to some incident. One can also face liquidity risk if an error causes major material damage, or regulatory risk if the system turns out to operate in breach of the rules.

5 RISK MANAGEMENT SOLUTIONS

Data protection. Both sides of AI applications require robust data protection. On the one hand, learning databases must be carefully screened to avoid violation of copyright, on the other hand, the nature of the information to be released on the output side may differ across users or fields of usage (NIST, 2024).

Rules of access. Access to certain AI-based systems must be limited to reduce the risk of different attacks (NIST, 2024).

Apply varied and representative learning data sets. Rather than relying on content picked up from any source indiscriminately, use well-screened data that present reality correctly. Algorithms recognising bias and ethical supervision should also be introduced.

Extensive tests and uninterrupted monitoring. It seems the AI systems of today cannot be applied without uninterrupted performance monitoring. It is the only way to notice if preliminary tests have failed to reveal some error or if a new er-

ror has appeared because either AI or reality have changed, or the error tolerance mechanisms applied are not perfect. Testing cannot be the responsibility of AI producers or user companies alone: governments and regulatory bodies must also take part since the free AI solutions that are available on the internet will be used by organisations that do not possess the skills needed for control. Micro and small enterprises can be so affected in Hungary. Such players can cause major economic damage through supplier chains and the part they play in employment unless some major error of freely available AIs is revealed in time.

Regulation. Proper legislation, ethical guidelines, and continued accountability are necessary so that AI development is given proper attention and targets the right goals. General requirements related to AIs must be made public in the same way as with vehicles so that the producers of AIs can build social expectations into their development processes and tests. In this regard, cross-border or even global collaboration of the regulatory bodies is especially important since a wide range of AI-based solutions of different quality are available on the global internet.

In addition, AI solutions – in contrast to traditional IT systems – learn and develop constantly. This can lead to unpredictable behaviour or unintended outcomes unless the systems are properly supervised and controlled. Obviously, setting up a proper regulatory framework at a certain point in time will not be enough; on the contrary, it will have to be continually adjusted to changes (European Commission, 2025).

Further, AI-related operational risks are manyfold. They often go beyond corporate operations since – in addition to technological errors – they involve strategic, reputational, legal, healthcare-related, psychological, and socio-economic challenges. Such risks may disrupt business processes, undermine customers' trust, and result in major financial or legal consequences (Cummings, 2024).

Staff training. Employers need to improve their employees' AI-related skills to mitigate risks resulting from improper use. Training is also necessary so that the staff that performed a task earlier can evaluate whether the results generated by AI are acceptable and judge whether true facts have been used or hallucinations have been eliminated.

Psychological awareness of staff. Organisations must pay special attention to the psychological challenges the staff may face because of the appearance of AI-based systems and daily work with them. They have to establish a supportive workplace environment to mitigate negative effects. You should also prepare in the long run for a situation when – 15 or 20 years from now – a generation will appear on the labour market for whom the use of AI tools will be second nature while they may not be able to apply AI-less methods.

Table 4
Risks and management methods of AI applications

Risk category	Risk type	Risk description	Best management methods
(I) Technological risks	Discrimination and toxicity	Bias and systemic distortions in AI decision-making	Use of representative and impartial datasets, ethical control
	(II) Organisational risks	Leakage of personal data, data theft and legal compliance	Effective data protection measures, regulation of access, encryption
	(III) Social and ethical risks	AI systems used for cyber-attacks, fraud, and abuse	Cyber security measures, fraud prevention technologies implemented
	Security and faults of AI systems	Unreliable systems, lack of transparency and security risks.	Ongoing testing, transparency mechanisms, compliance with regulations
(II) Organisational risks	Disinformation or misleading information	AI-generated false or misleading information spreads	Ongoing monitoring, integration of reliable sources, user training
	Human-computer interaction	Human independence diminishes, staff over-dependent on AI, technostress	Workplace training, ethical and psychological support, human supervision
(III) Social and ethical risks	Strategic and regulatory challenges	Deficiencies of regulations, responsibility issues and strategic challenges	Harmonisation of international regulations, definition of spheres of responsibility
	Socio-economic and environmental damage	AI impact on the workplace and society, increasing inequalities	Accommodation strategies at the workplace, social responsibility

Source: NFSZ, own design

*Supplements to organisational regulations and Codes of Ethics***.** Codes of Ethics already available in many places as well as internal organisational and procedural instructions need to be updated to include AI-related rules even if a company does not apply AI-based solutions yet. Experience has shown that to facilitate their work, some employees take the initiative and start using free AIs out of curiosity. It can involve major risks if proper regulations or training are lacking.

Incident management plans. As with any type of risk, incident management plans should be in place for AI-related risks; the relevant training must be arranged so that organisations are prepared if AIs become suddenly useless because of some fault or third-party attack. Therefore, no undertaking should allow a total collapse of human expertise needed to conduct activities taken over by AI in its organisation. It would be more reasonable for them to employ a lower number of

staff of high levels of specific skills to replace those with low or medium-level skills whose jobs have already been taken over by AI. In that way, they could supervise AI and could prepare for a potential AI loss.

Table 4 displays the best risk management practices for the different risk categories according to the professional literature. However, several paradoxes hamper their implementation.

6 ARTIFICIAL INTELLIGENCE PARADOXES

Risk management related to AI applications is made complicated because of the phenomena termed AI paradoxes in the professional literature. They are controversial areas in the development and use of AI. You can find an excellent summary of their background in Bakonyi's work (2024), presenting seven paradoxes, which was the primary source of this chapter. Based on similar research, Jazairy et al., (2024) have identified twelve paradoxes in corporate planning, while other authors focused on one or another paradox in their studies.

1.) *AI Stability Paradox*. It states that AI systems, particularly neural networks, were initially developed to provide a stable, high-accuracy description of different processes. However, it is clear that you cannot build such systems for certain problems (University of Cambridge, 2022), especially if the phenomenon or problem intended to be described changes over time.

2.) *Generative AI Paradox*. It states AIs can produce expert-level (or seemingly so) content while they have no real knowledge or understanding of the given issue. Thus, content can be persuasive even if it is false. It is dangerous because AI pretends to be an expert, which may mislead people with limited knowledge, particularly because (a) for people, understanding precedes the ability to prepare expert-level content, and (b) traditional IT tools are usually accurate and rely on facts and details (West–Aydin, 2024.) In addition, (c) confidence similar to that of AI is only typical of people who can recognise in-depth interrelations.

3.) *AI Trust Paradox*. It emphasises that technology acceptance and trust in it do not go hand in hand. Observations have proven that many people use AI-based systems but (correctly) do not trust them. Thus, they will question content generated with AI help even if they do not have any reason to do so because human authors check every fact to be properly proficient.

4.) *Domain expert paradox*. It is closely related to the trust and generative paradoxes. It describes the phenomenon that people have more confidence in algorithms developed with expert participation. On the other hand, experts

have no vested interest in participating as they have to work for their own substitution in the end (Jazairy et al., 2024). In addition, because they are experts, they are the ones to discover the errors of AIs, so they trust them the least. Also, while the correction of the errors is the least of their interests, they are the most suitable to carry them out.

5.) *Knowledge substitution paradox.* AI can substitute a certain level of organisational knowledge in some fields, but you need a higher level of knowledge than the one substituted to be able to offer a professional check of results generated by AI. What is more, lacking a real understanding of logic and context, AI can only rely on historical data and can hardly respond well to a new situation. Therefore, higher savings will not necessarily be achieved in an organisation if AI-based solutions are implemented in quality-sensitive fields. It is particularly true if a given task must be conducted even if AI is potentially not available. Jazairy et al. (2024) have pointed out if an organisation wants to introduce AI-based solutions, they must decide on how to have access to and use in future the special skills accumulated earlier in their employees' heads or in the corporate knowledge base. It is also essential that AI operated by a third party should only have limited access to company-specific knowledge such as partners' or employees' particulars or internal regulations.

6.) *Creativity Paradox.* The use of AI-based tools can significantly improve the creativity of low or medium-skilled employees, which is a great help provided the given skills have never existed in the organisation. However, if all competitors start using those tools, the advantage disappears, and originality turns into mass production since the "creative" contents are not new, but they are simple iterations of things made by others in the past (Osadcharya et al., 2024). There comes a time when you will again need high-level human knowledge for business success.

7.) *Task Substitution Paradox.* AIs are, in theory, applied to make employees' work easier and faster. However, it is widely believed employers will use worktime so liberated to reduce their workforce. So, the remaining staff will not work less just in a different way, while savings will improve corporate profits rather than working conditions. Still, it can happen that price competition on the distribution side swallows higher profits. Workers who have kept their jobs can buy goods cheaper. Meanwhile, the labour released can push wages down. In the end, nobody wins through the application of AIs. Others argue (Ferraro et al., 2024) that AI also creates new jobs, in other words, the technology is destructive and creative at the same time.

8.) *Time paradox*. While AIs promise to shorten the time needed to carry out different tasks, the implementation, fine-tuning, and training of artificial

intelligence are quite time-consuming, similar to traditional IT systems. In contrast to classical IT systems, the ongoing supervision of the operation and output of AI will always be an additional task (Osadcharya et al., 2024). Real savings can only be hoped for in the medium or long run.

9.) *Error Paradox.* Traditional IT solutions are typically more accurate than human work and well-implemented AI solutions can improve accuracy further. On the other hand, human errors are smaller but more frequent while AI is prone to less frequent but much bigger mistakes. So, risk management must be prepared to face fewer great-effect events rather than frequent low-effect events, which might be much more difficult. In addition, an organisation will experience those errors differently: people are willing to be more permissive with each other's mistakes than towards IT systems. Scaringi et al. (2024), for instance, present a case when physicians at a clinic assessed an algorithm developed to diagnose large vessel occlusion (LVO) unreliable because it gave a false positive signal in one case, although it performed very well otherwise.

10.) *Reference Paradox.* People are more inclined to believe predictions and assessments that are just slightly different from their own expectations. This can lead them to assess algorithms providing outcomes similar to human experts' expectations better when setting the parameters of AI applications. If, however, AI results are largely similar, doubts may arise as to their utility.

11.) *Experience Paradox.* AI solutions are good at recognising the patterns of human learning and making judgement on the basis of figures only. Human experts, however, often also consider qualitative aspects and are inclined to trust everyday experience and intuition that can be described verbally instead of complex mathematical models that are difficult to comprehend (Jazairy et al., 2024).

Distrust can be mitigated by presenting the surprising quantitative relations revealed by AI, but if they contradict human experience, the suspicion may arise that some specific past event has distorted the pattern. To accept AI logic by a wide group, experts are required to come across many examples when machine results have proved to be correct in the end. It is, however, time-consuming. It can, in fact, occur if an organisation bypasses the reference paradox by applying AI side-by-side with the earlier experts rather than in their place.

12.) *AI Alignment Paradox.* It means the closer AI gets to human perception and values the more vulnerable it will be to malignant influences (West–Aydin, 2024). That cannot be the goal, so it would be a grave mistake to make AI "fully human," although it could considerably improve its acceptance.

13.) *Superiority Paradox.* People working with AI may feel superior and inferior at the same time (Osadcharya et al., 2024). While they lag behind in factual knowledge or speed and often make mistakes, they are far ahead in creativity and understanding relationships, and their self-assessment is more realistic.

14.) *Illusive Connection Paradox.* AI chatbots imitating real people create an illusion of personal contact. However, if people realise AI is not human, it can have an adverse effect (Ferraro et al., 2024).

15.) *Satisfaction Paradox.* AI-based customer services work faster and provide more accurate information than people, which may increase customer satisfaction. On the other hand, in specific cases requiring empathy, the customers involved may be even less satisfied when a machine cannot help them, but they cannot reach a living person (Ferraro et al., 2024).

Jazairy et al., (2024) also present some problems that are dilemmas rather than paradoxes. Such a dilemma, for instance, is whether a company will fare better in terms of market advantage if it uses faster and more transparent traditional tools and waits until better AI tools are developed, or if they are pioneers in adapting systems that are not yet highly accurate.

You must clarify if you want to be reactive or proactive in planning. In other words, what is the reality of, and the costs involved if ad-hoc interventions are made, or what damage can planning mistakes cause. When AI systems are implemented, you must decide if you want a centralised or a decentralised system; if implementation is made step by step or simultaneously, and how much you are willing to reschedule already existing processes to fit AI.

HR must make a strategic decision on whether they want to employ traditional experts who also have AI skills in future or if they mainly want generalist AI gurus (Jazairy et al., 2024). The real dilemma, however, is whether you want to integrate AI tools into an existing organisational framework or if you should arrange all operations and jobs around AI.

The paradoxes revealed underline that the implementation of AI into an organisation is not simply a technological but also a strategic and philosophical challenge. AI can improve effectiveness and generate new problems at the same time while it can supplement or fully substitute human work. Being aware of the above paradoxes is key if you want to set out fundamental strategies. This means that companies need to manage AI not as a simple tool but as a factor that can fundamentally reshape their operating environment; therefore, it must be supervised all the time and regulated adaptively.

7 KEY FINDINGS

To sum up, the challenges involved in implementing and running AI-based solutions differ greatly from those you have encountered with traditional IT solutions. AI-based systems can produce content similar to that produced by experts without real understanding. Their assertiveness can easily mislead users accustomed to the accuracy of traditional IT systems. Thus, they need to be monitored constantly, and the results must be reviewed in detail. In addition, using them can really improve effectiveness up to a medium level of expertise only; going beyond that may result in mass-type products as creativity disappears.

Since AI-based solutions may offer novel targets for attack, while their operations are difficult to understand or can only be understood to a limited extent using traditional means, it is key that development is controlled through ethical and legal norms continually adapted to market changes; responsibilities should be clearly defined and systems already running should be monitored all the time.

Although artificial intelligence (AI) has become an integral part of modern workplaces and brought about major changes in business efficiency, innovation, and decision-making, the technology is still in its infancy. Its application carries complex operational risks that can appear in technical, ethical, legal, and socio-economic dimensions. To mitigate them, you must disseminate AI-related information even to organisations that do not officially operate such systems.

As AI technologies improve and spread, robust, ethical, and adaptive risk management gains importance. Organisations that manage AI-related risks proactively, learn continuously and are committed to ethical governance may achieve a competitive advantage while safeguarding sustainable operations and social responsibility.

Managing AI-related operational risks is not simply a technical or regulatory challenge but also a strategic exercise. Undertakings must implement adaptive risk management models, set out AI-specific codes of conduct, and have continuous monitoring. Regulatory authorities, on the other hand, should establish a framework to stimulate innovation and minimise the chances of abuse.

To sum up, undertakings should (1) implement transparency mechanisms that allow the supervision and traceability of decisions made by AI; (2) they should set up proactive risk management systems that allow ongoing assessment of the operation of AI, and (3) launch staff training and adopt ethical rules so that workers could use AI-based tools knowingly and safely.

Regulators have to (a) set up a uniform legal framework that clearly defines the responsibilities of the developers and users of AI; (b) international collaboration is key for that since AIs exert global impact so regulating them must also be global

and coordinated to eliminate loopholes. Finally, (c) effective supervisory mechanisms must be established to ensure ongoing supervisory monitoring of AI-based systems.

REFERENCES

AI Now Institute (2019): *AI Now Report 2019*. https://ainowinstitute.org/wp-content/uploads/2023/04/AI_Now_2019_Report.pdf (downloaded: 07.02.2025).

Bakonyi, Z. (2024): How can companies handle paradoxes to enhance trust in artificial intelligence solutions? A qualitative research. *Journal of Organizational Change Management*, 37(7), 1405–1426. https://doi.org/10.1108/JOCM-01-2023-0026.

Brynjolfsson, E. – McAfee, A. (2017): The business of artificial intelligence. *Harvard Business Review*. https://hbr.org/2017/07/the-business-of-artificial-intelligence (downloaded: 07.02.2025).

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 Global Institute*. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy (downloaded: 06.02.2025).

Buolamwini, J. – Gebru, T. (2018): Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 77–91.

CBSNews (2025): What is DeepSeek, and why is it causing Nvidia and other stocks to slump? https://www.cbsnews.com/news/what-is-deepseek-ai-china-stock-nvidia-nvda-asml/ (downloaded: 06.02.2025).

Cogent Infotech (2024): AI Risks: How Businesses Can Safeguard Their Future. https://www.cogentinfo.com/resources/ai-risks-how-businesses-can-safeguard-their-future (downloaded: 09.02.2025).

COSO (2017): Enterprise Risk Management: Integrating with Strategy and Performance. https://www.coso.org/_files/ugd/3059fc_61ea5985b03c4293960642fdce408eaa.pdf (downloaded: 07.02.2025).

Cummings, M. L. (2024): A Taxonomy for AI Hazard Analysis. *Journal of Cognitive Engineering and Decision Making*, 18(4), 327–332. https://doi.org/10.1177/15553434231224096.

Csáki, Cs. (2023): A mesterséges intelligencia elterjedéséből adódó kockázatok szisztematikus vizsgálata. In: Kovács, Z. (ed., 2023): A mesterséges intelligencia és egyéb felforgató technológiák hatásainak átfogó vizsgálata. Budapest, *Katonai Nemzetbiztonsági Szolgálat*, 27–50.

Dastin, J. (2018): Insight – Amazon scraps secret AI recruiting tool that showed bias against women. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/ (downloaded: 08.02.2025).

Davenport, T. – Ronanki, R. (2018): Artificial Intelligence for the Real World. *Harvard Business Review,* 96(1), 108–116. https://hbr.org/2018/01/artificial-intelligence-for-the-real-world.

Devlin, J. – Chang, M. W. – Lee, K. – Toutanova, K. (2018): BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*:1810.04805. https://arxiv.org/abs/1810.04805.

Domokos, A. – Sajtos, P. (2024): Mesterséges intelligencia a pénzügyi szektorban – Innováció és kockázatok. *Hitelintézeti Szemle*, 23(1), 155–166.

Egan, M. (2024): AI is replacing human tasks faster than you think. https://edition.cnn.com/2024/06/20/business/ai-jobs-workers-replacing/index.html (downloaded: 09.02.2025).

Európai Bizottság (2025): A mesterséges intelligenciáról szóló rendelet. https://www.consilium.europa.eu/hu/policies/artificial-intelligence/ (downloaded: 07.02.2025).

Ferraro, C. – Demsar, V. – Sands, S. – Restrepo, M. – Campbell, C. (2024): The paradoxes of generative AI-enabled customer service: A guide for managers. *Business Horizons*, 67(5), 549–559. https://doi.org/10.1016/j.bushor.2024.04.013.

Financial Times (2024): Employers look to AI tools to plug skills gap and retain staff. https://www.ft.com/content/9cf58a76-5245-4cdf-9449-239e90077eb5 (downloaded: 07.02.2025).

Hassel, A. – Özkiziltan, D. (2023): Governing the work-related risks of AI: implications for the German government and trade unions. *Transfer: European Review of Labour and Research*, 29(1), 71–86. https://doi.org/10.1177/10242589221147228.

Hill, K. (2021): The Secretive Company That Might End Privacy as We Know It. *The New York Times*. https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html (downloaded: 09.02.2025).

Hill, K. (2024): Clearview AI Used Your Face. Now You May Get a Stake in the Company. *The New York Times*. https://www.nytimes.com/2024/06/13/business/clearview-ai-facial-recognition-settlement.html (downloaded: 09.02.2025).

Holweg, M. – Younger, R. – Wen, Y. (2022): The reputational risks of AI. *California Management Review*. https://cmr.berkeley.edu/2022/01/the-reputational-risks-of-ai/ (downloaded: 08.02.2025).

Jazairy, A. – Shurrab, H. – Chedid, F. (2024): Impact pathways: walking a tightrope—unveiling the paradoxes of adopting artificial intelligence (AI) in sales and operations planning. *International Journal of Operations – Production Management*, 45(13), 1–27. https://doi.org/10.1108/IJOPM-07-2024-0582.

Kreps, S. – George, J. – Lushenko, P. – Rao, A. (2023): Exploring the Artificial Intelligence ‚Trust Paradox': Evidence from a Survey Experiment in the United States. *PLOS ONE*, 18(1), e0288109. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0288109 (downloaded: 09.02.2025).

Krisztián, B. – Nemeskéri, Z. (2014): A Taylori elvek a magyar gazdaságban. *Taylor Gazdálkodási és Szervezéstudományi Folyóirat*, 6(1-2), 498–508., https://ojs.bibl.u-szeged.hu/index.php/taylor/article/view/12838 (downloaded: 19.02.2025).

Krizhevsky, A. – Sutskever, I. – Hinton, G. (2012): ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 25. https://doi.org/10.1145/3065386.

Lestari, N. S. – Rosman, D. – Veithzal, A. P. – Zainal, V. R. – Triana, I. (2023): Analysing the Impact of Robot, Artificial Intelligence, and Service Automation Awareness, Technostress and Technology Anxiety on Employees's Job Performance in The Foodservice Industry, *2023 5th International Conference on Cybernetics and Intelligent System (ICORIS)*, Pangkalpinang, Indonesia, 2023, pp. 1-6, doi: 10.1109/ICORIS60118.2023.10352286.

Majumder, S. – Yashraj, A. (2024): Mitigating AI-Driven Flash Crashes. http://dx.doi.org/10.2139/ssrn.4950688 (downloaded: 08.02.2025).

McKinney, S. M. – Sieniek, M. – Godbole, V., et al. (2020): International evaluation of an AI system for breast cancer screening. *Nature, 577*(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6.

MIT (2024): The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence, https://arxiv.org/pdf/2408.12622 (downloaded: 2025.02.07.).

NIST (2024): Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, https://doi.org/10.6028/NIST.AI.600-1 (downloaded: 07.02.2025).

NTSB (2019): Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018, Accident Report https://www.ntsb.gov/investigations/accidentreports/reports/har1903.pdf (downloaded: 08.02.2025).

Osadchaya, E., Marder, B. – Yule, J. A. – Yau, A. – Lavertu, L. – Stylos, N. – Oliver, S. – Angell, R. – Regt, A. de – Gao, L. – Qi, K. – Zhang, W. Z. – Zhang, Y. – Li, J. – AlRabiah, S. (2024): To ChatGPT, or not to ChatGPT: Navigating the paradoxes of generative AI in the advertising industry. *Business Horizons*, *67*(5), 571–581. https://doi.org/10.1016/j.bushor.2024.05.002.

Pearson, J. – Zinets, N. (2022): Deepfake footage purports to show Ukrainian president capitulating. https://www.reuters.com/world/europe/deepfake-footage-purports-show-ukrainian-president-capitulating-2022-03-16/ (downloaded: 09.02.2025).

Ragu-Nathan, T. S. – Tarafdar, M. – Ragu-Nathan, B. S. – Tu, Q. (2008): The Consequences of Technostress for End Users in Organizations: Conceptual Development and Empirical Validation. *Information Systems Research*, 19(4), 417–433.

Reuters (2025): DeepSeek gives Europe's tech firms a chance to catch up in global AI race. https://www.reuters.com/technology/artificial-intelligence/deepseek-gives-europes-tech-firms-chance-catch-up-global-ai-race-2025-02-03/?utm_source=chatgpt.com (downloaded: 06.02.2025).

Reuters (2025): Ride-hailing platform Lyft ties up with Anthropic for AI-powered customer care. https://www.reuters.com/technology/artificial-intelligence/ride-hailing-platform-lyft-ties-up-with-anthropic-ai-powered-customer-care-2025-02-06 (downloaded: 07.02.2025).

Ringly (2025): 10 Real-Life Examples of Artificial Intelligence in 2025. https://www.ringly.io/blog/10-examples-of-artificial-intelligence-in-2025 (downloaded: 07.02.2025).

Scaringi, J. A. – Mctaggart, R. A. – Alvin, M. D. – Atalay, M. – Bernstein, M. H. – Jayaraman, M. V. – Jindal, G. – Movson, J. S. – Swenson, D. W. – Baird, G. L. (2024): Implementing an AI algorithm in the clinical setting: a case study for the accuracy paradox. *European Radiology*. https://doi.org/10.1007/s00330-024-11332-z.

Shepardson, D. – Jin, H. (2021): U.S. opens probe into Tesla's Autopilot over emergency vehicle crashes, https://www.reuters.com/business/autos-transportation/us-opens-formal-safety-probe-into-tesla-autopilot-crashes-2021-08-16/ (downloaded: 09.02.2025).

Szabó, H. (2023): A mesterséges intelligencia biztonsági kockázatai egy új korszak kezdetén. *Nemzetbiztonsági Szemle*, 11. évfolyam (2023) 4. szám 35–46., https://doi.org/10.32561/nsz.2023.4.3.

University of Cambridge (2022): Mathematical Paradox Demonstrates the Limits of AI. *University of Cambridge*. https://www.cam.ac.uk/research/news/mathematical-paradox-demonstrates-the-limits-of-ai (downloaded: 09.02.2025).

Walz, E. (2024): NHTSA opens safety probe for up to 2.4M Tesla vehicles, https://www.automotive-dive.com/news/nhtsa-opens-investigation-tesla-fsd-odi-crashes-autopilot/730353/ (downloaded: 09.02.2025).

West, R. – Aydin, R. (2024): The Generative AI Paradox: ‚What It Can Create, It May Not Understand'. *arXiv preprint arXiv*:2311.00059. https://arxiv.org/abs/2311.00059 (downloaded: 09.02.2025).

West, R. – Aydin, R. (2024): There and Back Again: The AI Alignment Paradox. *arXiv preprint arXiv:*2405.20806. https://arxiv.org/abs/2405.20806 (downloaded: 09.02.2025).

Yin, Z. – Wang, H. – Horio, K. – Kawahara, D. – Sekine, S. (2024): Should We Respect LLMs? A Cross-Lingual Study on the Influence of Prompt Politeness on LLM Performance. *arXiv:*2402.14531. https://arxiv.org/abs/2402.14531.

In the article the author used ChatGPT 40 and Perplexity.AI Pro to generate ideas and research sources. The author did not use texts generated by AI-based tools directly; he reviewed all facts and references separately.