WASD
WORLD ASSOCIATION FOR SUSTAINABLE DEVELOPMENT

6

OUTLOOK 2025

BALANCING AI SPEED AND CONTROL

PATRICK ZALDA
Research Student
DBA Programme
Heriot-Watt University, UK
Email: pz2001@hw.ac.uk
ORCID: 0009-0000-0053-8116

ABSTRACT

PURPOSE: There are many reasons for the existence and further development of Artificial Intelligence (AI). However, with the rapid technical development of AI, 'guard rails' are needed to ensure its safe and human-centred use.

DESIGN/METHODOLOGY/APPROACH: The research conducts a Delphi study in which experts from professional, academic and political environments are interviewed in iterative surveys. The aim of the study is to determine to what extent ethical aspects can be integrated into the implementation process of AI from the perspective of experts in AI.

FINDINGS: The initial suggestions from the research are that organisations are not well aware of the ethical issues themselves, and even where some do have a strong awareness of such issues, they have not really considered how to take ethical aspects of AI into account.

ADDED VALUE/PRACTICAL IMPLICATIONS: Guidelines for the safe and human-centred use of AI are new. Companies that use or produce AI will be able to benefit from the full research results, especially in the European Union (EU) where the AI Act is already in force; compliance with this Act will be helped by several strands of this ongoing research.

KEYWORDS: *Artificial Intelligence; Ethics; Delphi Study; Risk Management; Governance; EU AI Act.*

CITATION: Zalda, P. (2025): Balancing AI Speed and Control. In Ahmed, A. (Ed.): *United Nations: What Next After 2030 Agenda and SDGs.* World Sustainable Development Outlook 2025, Vol. 21, pp.83-96. WASD: London, United Kingdom.

RECEIVED: 14 October 2025 / **REVISED:** 12 November 2025 / **ACCEPTED:** 24 November 2025 / **PUBLISHED:** 30 December 2025

DOI: 10.47556/B.OUTLOOK2025.23.7

INTRODUCTION

Artificial intelligence (AI) is everywhere, and the use of generative AI has made it very easy for individuals to produce texts, images and sounds, just to name a few examples. The Chief Executive Officer (CEO) of Alphabet compares the current situation as ‘AI will have a more profound impact on humanity than fire, electricity and the internet’ (Prakash, 2023). However, it is not just private individuals who use AI. More and more companies have varying degrees of AI in use or are planning to use AI in their organisation. This rapid technological progress could lead to societies losing trust in AI if its potential risks are not addressed. It is therefore important to determine how ethical aspects can be incorporated into the AI implementation process so that AI can be used in a human-centred and transparent manner.

LITERATURE REVIEW

‘Creative Destruction’ and AI – Does Schumpeter’s Theory Apply?

The introduction of machines to replace manual jobs, such as during industrialisation, resulted in some people becoming unemployed. However, new work was also created.

This process was defined by Joseph Schumpeter in the 1940s as ‘Creative Destruction’ and says nothing more than that current market situations and conditions are replaced by newer technologies, resulting in social and economic growth. Schumpeter’s theory is based on economic change driven by innovation to develop new consumer goods. These goods can have an impact on existing product lines, the logistics of goods, organisations, or markets. However, the mere innovation of goods does not explain Schumpeter’s theory. Rather, it means that an innovation opens a completely new market and that new products and structures are created based on this new market (Bommineni, 2021, p.14f). Schumpeter (2003, p.83) sees capitalism as the starting point for constantly changing circumstances and includes the entrepreneur as a variable in his model. As an example, Schumpeter (2003) cites, among other things, the process of agriculture; this was initially carried out by purely physical labour and was subsequently industrialised by agricultural machinery. A process of market displacement created by capitalism thus resulted in ‘creative destruction’ (Schumpeter, 2003, p.85). Overall, Schumpeter’s model consists of three characteristics: (1) growth is innovation-driven, which leads to new and better products; (2) innovations generate profit in monopoly-like structures (‘monopoly rents’); (3) ‘creative destruction’ is unstoppable when legacy technologies are replaced by innovations and new companies enter the market as a result (Dosi *et al.*, 2016, p.64).

 DOI: 10.47556/B.OUTLOOK2025.23.7

However, Schumpeter's model can be criticised for the fact that innovation-driven growth has both short-term and long-term effects and such a distinction is not taken into account. Specifically, this refers to the distinction between 'business cycles' (for example, according to Keynes) and so-called 'long waves' (for example, according to Kondratieff).

Bommineni (2021) describes that in a 'creative destruction' there are challenges on the legal, political and ethical level. This could mean, for example, that people do not trust a new technology because it changes their usual production processes and therefore saves on labour. These societal concerns are understandable and, in this respect, causal, because new markets and market participants can arise through 'creative destruction'. We cannot say at this stage whether AI is considered a basic innovation, but what we do know is that security and regulation will be needed if AI as a basic innovation leads to a sixth wave. Regulatory means laws that protect a society: in the case of AI, this could be the definition of responsibilities in the event of social damage.

Industrialisation can be used to illustrate ex post what a basic innovation ('steam powered mechanisation') led to both socially and above all politically: new laws were passed to protect society. In 1802, the Health and Morals Act was passed in Great Britain; this Act restricted child labour in factories. This prevented night shifts and daily working hours beyond 12 hours and provided basic education for children (UK Parliament, 2022).

These types of regulations were enacted by governments in many countries in order to minimise risks to social co-existence. Today, a trend can be identified that is not significantly different from 200 years ago: regulations are enacted to protect society. This approach creates social trust in technologies such as AI. Despite regulations, however, there is a risk that mistakes are made on the basis of political decisions that, contrary to expectations, do not protect society and thus a learning effect from historical mistakes fails to materialise.

Should an AI be regulated too strongly and thus meet Schumpeter's definition of a 'basic innovation', then the question arises, with regard to the above examples, whether political decisions and regulations could once again lead to society not learning from its mistakes.

Placing Digital Ethics in the Context of Risk Management

There are many examples of AI use cases that allow AI to be used either positively or negatively. Positive areas of usage, in the sense of Floridi *et al.* (2020) and the 'AI for Social Good' (AI4SG) model, could be seen, for example, in the banking industry in fraud detection, in customer service using chatbots or in the recruitment process. Negative areas

DOI: 10.47556/B.OUTLOOK2025.23.7

of usage would be, for example, when AI is used for warfare, creating deep fakes, or when the use of AI leads to negative societal impacts. All AI use cases have one issue in common: AI is associated with risks that require consideration in the corporate context. A use case for AI that is positive at first glance can become a significant risk over time and, for example, significantly damage a company's reputation.

The need to establish AI risk management is seen by governments, the scientific world and now by civil society, especially when it comes to sensitive issues such as the use of GenAI and the associated 'job reduction of creative jobs' (Greenhouse, 2023). Due to certain risk potentials of AI applications, the AI Act of the European Union has defined the materiality levels 'unacceptable risk', 'high risk', 'limited risk', 'low and minimal risk' (EPRS, 2022, p.5). To simplify the situation, the terms 'strong' and 'weak' AI are used. For strong AI ('unacceptable and high risk'), the AI Act specifically defines the requirement that a Risk Management System (RMS) must be in place for those companies that produce or use AI (EC, 2021, Chapter 2, Article 9). The minimum components of an RMS are the identification and analysis, assessment, and treatment of 'high risk' AI. The National Artificial Intelligence Initiative Act of 2020 in the United States also requires that a risk management framework be regularly updated (Thornberry, 2020, p.1179). In particular, risks should be evaluated, and risk mitigation measures described to ensure that AI is secure and robust against malicious attacks and that these testing mechanisms lead to more transparency, verifiability and security.

Such an RMS could consist of the following stages: keep an inventory of AI, conduct compliance assessments, establish an AI governance system. This AI risk management could use the AI Act as a framework, for example, so that the regulatory requirements are defined internally within the company. Afterwards, AI applications would be inventoried to subsequently carry out a conformity analysis (requirements according to the AI Act vs actual state in the company). In the last step, an AI governance system is established (Benjamin *et al.*, 2021).

However, an RMS provides for responsibilities to be concretely defined. Here, though, the following ethical question arises: who is responsible for not only the successes, but also for the harm caused by AI applications (Coeckelbergh, 2020, p.2052)? Coeckelbergh (2020, pp.2053-2058) makes clear how challenging the question is to answer by asking whether technology can be held responsible for an outcome at all. The company IBM (IBM, 2022) tends to see the responsibility of AI outcomes being with AI designers and developers; scientists, such as Arrieta *et al.* (2020), tend to avoid the question and do not give any specific clues to answer the 'accountability' question. A discussion on a meta-level will not be pursued, because no generally valid result would be reached. However, a result must be reached if AI is used or produced in the company, because otherwise an RMS would not be efficient and effective.

 DOI: 10.47556/B.OUTLOOK2025.23.7

Therefore, the following pragmatic procedure could be applied: if an AI is used, for example, for the recruitment process, then the head of the HR department is responsible for the results, whether positive or negative. This means that the 'place of use' of an AI should provide information on responsibility. This incorporation makes sense because risks are always process-related. If an AI risk is identified, then it must be assigned to a corresponding process and thus to a process owner. This assignment could clarify the question of responsibility, at least in the company context.

Nevertheless, there is another challenge: how to identify, analyse and assess AI risks. To do this, the AI must be understood and the so-called 'black-box' problem (Coeckelbergh, 2020, p.2060) must be avoided. This means that the result of a technical function of an AI must be explainable and interpretable (Sarker, 2021, p.420). If a technical function is not explainable and interpretable, then in terms of qualitative risk assessment it is not possible to assess what consequences an AI may have on a company's reputation, resources, or strategy. Similarly, quantitative risk assessment, i.e., what financial impact a risk can have, is made more difficult by a 'black-box' problem, because assumptions and probabilities of occurrence for a risk cannot be clearly defined.

But how do these risks relate to digital ethics? The following definition could be given for digital ethics in a business context: how do I (as a company) behave well under the conditions of digitalisation? 'Good' can mean, on the one hand, that attention is paid to the ecological blueprint of AI applications, that AI is used responsibly, that no people are harmed, etc. On the other hand, 'good' can mean that a company does not harm others. Hagendorff (2020, p.108) describes the problem that, from a corporate perspective, digital ethics is currently only beneficial for public relations purposes, because taking ethical aspects into account in the development process of an AI initially costs both money and time. Nevertheless, business ethics is part of corporate governance with integrity and is relevant to risk management because reputational risks can have a major financial impact. All of the issues identified above will be deeply investigated as this research project goes forward using the Delphi method.

RESEARCH METHOD

The Delphi Method

This methodology follows a more qualitative and inductive research approach. The Delphi method conducts iterative interviews with experts, who are described as 'panellists' (Petry *et al.*, 2007). These 'panellists' are accompanied by a 'survey coordinator' (Hirschhorn, 2019, p.311) and pursue the goal of reaching a consensus among the participants on a certain topic. In this way, the current state of knowledge is questioned and supplemented

DOI: 10.47556/B.OUTLOOK2025.23.7

by mainly qualitative data, through an explorative approach (Iqbal and Pipon-Young, 2009). Hirschhorn (2019) describes the procedure in such a way that after an initial round of interviews, the participants receive feedback on their response options and in a second (third, fourth, ...) round they can reflect on their initial response to a question. In this iterative process, interviews and therefore possible answers are reflected upon until most participants have agreed on an answer and a consensus exists. According to Hirschhorn (2019), the Delphi method is very flexible in its usage and less likely to be static, such as fully structured interviews. For example, the survey co-ordinator has the option to differ in the procedural approach if a consensus is reached early in the interview rounds. In such a case, no further iterations are needed.

The procedural approach of the Delphi method would first include a pilot study to verify the quality of the survey (Iqbal and Pipon-Young, 2009). After successfully accepting and evaluating the pilot study, data collection would take place in the form of semi-structured interviews, which would be conducted in iterative processes. This means that a first round of interviews will take place with all experts. The answers of the first round are evaluated and enriched with anonymised feedback. In the second, third (...) round, the questionnaire is adapted in each case so that at the end of the interview rounds there is a consensus on the questions asked.

The described procedure has the advantage that in-depth knowledge can be generated. However, besides the advantages of a Delphi method, there are also challenges in the research design. Bernal *et al.* (2019) see the challenge that no immediate feedback is obtained on problems of understanding and that it is difficult to recruit experts to participate. Regardless of how participants are recruited, it cannot be excluded that participants might drop out during the empirical study and are no longer available for further rounds of questions. This would mean that a strategy would have to be developed in advance of the study to create incentives for participants to be available for the entire duration of the empirical study. Similarly, the research set-up seems very time-consuming because iterative rounds of questions must be conducted with X number of people before a consensus is reached.

Data Gathering and Sampling

Within the Delphi study, first quantitative and then qualitative data will be collected. Brady (2015, p.3) provides for a questionnaire with ‘semi-open’ questions, which was developed based on extensive literature research. Three interview rounds represent the common procedure in the Delphi methodology. The first round of interviews therefore contains the originally elaborated questionnaire. The second round of interviews uses

DOI: 10.47556/B.OUTLOOK2025.23.7

an adapted questionnaire that has incorporated feedback from the first round. The third round of interviews uses a questionnaire that emerges from the two previous interviews (Brady, 2015, p.3). If no consensus is found on a topic after three rounds, then the iterative procedure ends in this research.

In contrast to Petry *et al.* (2007, p.339), the objective in this work is not to achieve an 80% 'criterion of consensus', but a two-thirds majority, i.e., 66%. A two-thirds majority is seen as an appropriate limit for implementing majority decisions. This means that when a consensus is reached, the iterative procedure ends, and the participants are not consulted further.

The most important actors or participants in this work could be AI developers and users, researchers in the field of AI/ethics/philosophy, as well as legislators or people related to the AI Act. Therefore, expert knowledge is required to generate qualified data. Brady (2015, p.3) recommends that 'purposive sampling' should be used in the Delphi study. This means that participants are selected for a study because certain attributes are essential for the investigation. These attributes must be defined in advance so that appropriate experts can be recruited for the study.

More specifically, people are considered as experts for this study if they can be divided into the following categories: theory-experts, practice-experts, political-experts (following Petry *et al.*, 2007, p.334). A quota of 33% is targeted for each of these three categories; in total, this means that 30 people should take part in this Delphi study. Petry *et al.* (2007) interviewed 45 persons in their study, while Iqbal and Pipon-Young (2009) assume that 10-50 persons are appropriate for a Delphi study. Collins (2010, p. 43) also describes that a smaller sample size is more advantageous in inductive research because it allows in-depth knowledge to be generated. Therefore, a sample of 30 seems appropriate for this current research.

The following criteria will help to pre-select appropriate experts for this research:

- **Theory-Experts:** Researchers from the fields of robotics, technical sociology, digital ethics, technical philosophy, machine learning/artificial intelligence with at least one year of academic experience; participants in scientific conferences or working groups related to AI; willingness to participate in a maximum of three iterative interviews; the ability to express complicated content in a verbally comprehensible way.
- **Practice-Experts:** Persons related to AI in a professional environment, e.g., AI developers, IT architects, consultants, persons who use AI in a professional environment on a regular basis (at least three days a week); willingness to participate in a maximum of three iterative interviews; ability to express complicated content in a verbally comprehensible way.

DOI: 10.47556/B.OUTLOOK2025.23.7

- **Political-Experts:** Persons who work on AI regulation (e.g., AI Act) or have interfaces to the AI Act, such as non-governmental organisations (NGOs), employees of government-related institutions, employees of national and European institutions; willingness to participate in a maximum of three iterative interviews; the ability to express complicated content in a verbally comprehensible way.

RESULTS

This work follows an inductive research approach in which results from interviews and questionnaires were aggregated to extract data patterns; based on these, a consensus/dissent on suggested propositions were analysed. When analysing results, it is possible that correlations or patterns may be seen that make sense in the data. However, these results cannot be causal. Since this work is more qualitative in nature and does not use mathematical models, causality is not established through statistical methods or mathematical models, but rather through an understanding of processes, social mechanisms and structures. In other words: does the context of the interviews, the interpretation of the results and the progression of the results provide a causal narrative that explains certain results? The aim of this work is to find plausible explanations for certain results whose statements are supported by the data. It is not the aim to find absolute proof that, for example, compares and explains correlation between certain variables. The two research questions were formulated at the beginning of the thesis and will be answered by the analysed data.

Research Question 1: How can digital ethics be sustainably integrated into the development process of AI to ensure that AI is safe?

The long-term integration of digital ethics into the AI development process requires structural, technical, and social measures. These are clear from both the qualitative data from the first round and the quantitative data from the second round of the main study.

Early and systematic integration

Ethical aspects must be considered early in the development process, for example as non-functional requirements (NFRs). This ensures that ethics are not seen as a retrospective thought but are considered right from the start. These NFRs relate not only to the database to avoid bias or protect minorities, but also to technical functions to create transparency, traceability and robustness.

 DOI: 10.47556/B.OUTLOOK2025.23.7

Institutional measures

Ethical aspects can be integrated through checklists, certifications, ethics modules (for developers, students, professionals), ethics boards (in companies) and regular reviews of training data.

Interdisciplinary and diversity-sensitive teams

Taking different perspectives into account, e.g., through diverse development teams, is crucial for comprehensively understanding and implementing ethical requirements.

Orientation towards frameworks

The EU AI Act and other ethical frameworks are used as guidance, even if cultural differences make implementation difficult (for companies). The mere existence of frameworks could lead to greater security for people, even if implementation is challenging.

Consideration of context dependency

Ethics is considered particularly relevant for AI systems that interact directly with people, e.g., in medicine, human resources or automated decision-making. In technical or purely data-driven applications with no direct human impact, ethics is often considered less important. In addition, the social impact of AI determines whether and to what extent ethical requirements must be considered.

Consideration of implementation barriers

Even if ethics is recognised as important, there are specific barriers that make implementation difficult. For example, there is a lack of standardised procedures for implementing ethics technically, e.g., in algorithms or data structures. The operationalisation of ethical principles is also unclear, i.e., how to write and implement ‘fairness’ or ‘responsibility’ as code. In addition to technical barriers, there are cultural barriers. Different understandings of values in different countries or organisations lead to conflicts: what is considered ethically correct in one context may be considered as problematic in another. In addition, companies find themselves caught between ethics and efficiency, because taking ethics into account can lead to higher development costs (e.g., through additional testing processes, certifications, diverse teams). Another source of conflict can be a longer time-to-market, which is seen as a competitive disadvantage.

DOI: 10.47556/B.OUTLOOK2025.23.7

These six points indicate that digital ethics can be sustainably integrated into AI development if it is established early on, systematically and institutionally. Specific measures such as ethics checklists, certifications, interdisciplinary teams and alignment with legal frameworks are key. However, implementation must be context-specific and actively address technical and cultural barriers.

Research Question 2: Will AI be stronger if it is ethical?

The data show that ethics can significantly increase the strength of AI, not automatically, but under certain conditions – both socially and technically.

Strengthening through trust and acceptance

Ethics promotes social trust, fairness and acceptance. These are key factors for the broad use of AI. The existence of frameworks such as the EU AI Act alone leads to greater acceptance of AI because people feel more secure when it comes to ethical issues.

Improving technical quality

Ethics contributes to data quality, governance and robustness, e.g., through bias controls, transparent decision-making logic and active risk management. These aspects strengthen the reliability and security of AI systems.

Impact of ethics on AI performance

Ethics can specifically improve AI performance when used in areas where AI interacts with humans, decision-making or social responsibility. In such cases, ethics leads to greater acceptance among users and stakeholders or increases the trustworthiness of AI systems (e.g., in medical technology). The impact of ethics on AI is not universal but depends on the specific use case. In purely technical or internal applications without direct human impact, the impact of ethics on the strength of AI is considerably lower.

Restrictions due to economic and technical conflicts of interest

Ethics can restrict the development and performance of AI if it collides with other objectives. For example, additional costs of compliance, certifications and, in some cases, ethical audits may increase costs and extend development times. There are also technical conflicts of interest in terms of technical complexity, because ethical requirements cannot

 DOI: 10.47556/B.OUTLOOK2025.23.7

be automated and require manual evaluation by humans (e.g., when defining ‘fairness’ or ‘responsibility’). There is also a lack of clear technical specifications for implementation; this makes the integration of ethical principles difficult to implement in practice. These limitations and conflicts of interest result in ethics being seen as an obstacle rather than a strength in certain cases.

From the four points mentioned above, it can be concluded that AI can be stronger if it is ethically oriented. This is particularly true in terms of trust, technical quality, social acceptance and relevance. However, the effect of this strength is context-dependent and can be limited by economic or technical conflicts of interest. For this reason, **ethics can be seen as a stabilising factor rather than a universal amplifier.**

CONCLUSIONS

The development of AI has been very dynamic in recent years. Large language models have been established, many AI use cases have been developed, and companies, science and our society are benefiting from this technical development. The consideration of ethical aspects in this AI context is highly relevant, but limitations and further research needs can be considered in terms of both content and methodology:

- **Limitation 1 –** Technology is changing: As technological developments in the field of AI are dynamic, the results of this study must always be interpreted in the context of the respective state of technology at the time of the investigation.
- **Limitation 2 –** Classification of the term ‘ethics’: This study does not claim to take a position on the ‘right’ ethical perspective. It understands ethics as an open framework for discourse, the specific form of which requires further research.
- **Limitation 3 –** Availability of experts: The selection and number of experts interviewed was limited by their restricted availability in the field of research, which is why the range of perspectives is limited.
- **Limitation 4 –** Validity of the research results. The results of this deductive study are to be understood as derivations whose validity is based on methodological consistency and contextual classification.

In summary, this means that the increasing popularity of generative AI has fundamentally changed the framework conditions for scientific work and raises new questions about integrity, ethical responsibility and social impact. In particular, the potential simulation of ethical principles by AI, psychological projection on AI systems, and the conflicts between economic interests and ethical standards highlight the need for in-depth interdisciplinary research on security, trust, and ethical governance in the age of AI.

DOI: 10.47556/B.OUTLOOK2025.23.7

REFERENCES

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S. Gil-Lopez, S., Molina, D., Benjamins, R. Chatila, R. and Herrera, F. (2020): Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, Vol. 58, pp.82-115. Available at: https://doi.org/10.1016/j.inffus.2019.12.012

Benjamin, M., Buehler, K., Dooley, R. and Zipparo, P. (2021): What the draft European Union AI regulations mean for business. Available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/what-the-draft-european-union-ai-regulations-mean-for-business Accessed: 27 October 2022

Bernal, R., San-Jose, L. and Retolaza, J.L. (2019): Improvement Actions for a More Social and Sustainable Public Procurement: A Delphi Analysis. *Sustainability*, Vol. 11, No. 15, p.4069. Available at: https://www.mdpi.com/2071-1050/11/15/4069

Bommineni, S. (2021): Artificial Intelligence: Creative Destruction in the Autonomous Vehicles Industry. *International European Extended Enablement in Science, Engineering & Management (IEEE-SEM)*, Vol. 9, No. 10. Retrieved from: https://www.ieeesem.com/researchpaper/Artificial_Intelligence_Creative_Destruction_in_the_Autonomous_Vehicles_Industry.pdf

Brady, S.R. (2015): Utilizing and Adapting the Delphi Method for Use in Qualitative Research. *International Journal of Qualitative Methods*, Vol. 14, No. 5, p. 1609406915621381. Available at: https://doi.org/10.1177/1609406915621381

Coeckelbergh, M. (2020): Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, Vol. 26, No. 4, pp.2051-2068. Available at: https://doi.org/10.1007/s11948-019-00146-8

Collins, H. (2010): *Creative Research: The Theory and Practice of Research of the Creative Industries*. AVA Publishing. Available at: https://doi.org/10.5040/9781474247115

Dosi, G., Napoletano, M., Roventini, A. and Treibich, T. (2016): Micro and macro policies in the Keynes+Schumpeter evolutionary models. *Journal of Evolutionary Economics*. Vol. 27, No. 1, pp.63-90. Available at: https://doi.org/10.1007/s00191-016-0466-4

European Commission (EC) (2021): *Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN Accessed: 30 September 2022

European Parliamentary Research Service (EPRS) (2022): *Regulatory framework proposal on artificial intelligence*. Available at: https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

DOI: 10.47556/B.OUTLOOK2025.23.7

Floridi, L. Cowls, J., King, T.C. and Taddeo, M. (2020): How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics*, Vol. 26, pp.1771-1796. Available at: https://doi.org/10.1007/s11948-020-00213-5

Greenhouse, S. (2023): US experts warn AI likely to kill off jobs – and widen wealth inequality. *The Guardian*, 8 February. Available at: https://www.theguardian.com/technology/2023/feb/08/ai-chatgpt-jobs-economy-inequality Accessed: 21 April 2023.

Hagendorff, T. (2020): The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, Vol. 30, No. 1, pp.99-120. Available at: https://doi.org/10.1007/s11023-020-09517-8

Hirschhorn, F. (2019): Reflections on the application of the Delphi method: lessons from a case in public transport research. *International Journal of Social Research Methodology*, Vol. 22, No. 3, pp.309-322. Available at: https://www.tandfonline.com/doi/full/10.1080/13645579.2018.1543841

IBM (2022): Accountability. AI designers and developers are responsible for considering AI design, development, decision processes, and outcomes. Available at: https://www.ibm.com/design/ai/ethics/accountability/#:~:text=AI%20designers%20and%20developers%20are,objective%20system%20of%20logical%20decisions. Accessed: 26 April 2023

Iqbal, S. and Pipon-Young, L. (2009): *The Delphi Method*. The British Psychological Society. [Online] Available at: https://www.bps.org.uk/psychologist/delphi-method Accessed: 19 September 2023.

Petry, K., Maes, B. and Vlaskamp, C. (2007): Operationalizing quality of life for people with profound multiple disabilities: A Delphi study. *Journal of Intellectual Disability Research*, Vol. 51, No. 5, pp.334-349. Available at: https://doi.org/10.1111/j.1365-2788.2006.00882.x

Prakash, P. (2023): Alphabet CEO Sundar Pichai says that A.I. could be 'more profound' than both fire and electricity—but he's been saying the same thing for years. *Fortune Magazine*. Available at: https://fortune.com/2023/04/17/sundar-pichai-a-i-more-profound-than-fire-electricity/ Accessed: 28 December 2024.

Sarker, I.H. (2021): Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, Vol. 2, No. 6, pp.1-20. Available at: https://doi.org/10.1007/s42979-021-00815-1

Schumpeter, J.A. (2003): *Capitalism, Socialism and Democracy*. George Allen & Unwin (Publishers) Ltd 1976. London and New York, 2003.

Thornberry, W.M. (2020): National Defense Authorization Act for Fiscal Year 2021. U.S. Government Publishing Office, Washington 2020. Available at: https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210 Accessed: 24 September 2022.

DOI: 10.47556/B.OUTLOOK2025.23.7

UK Parliament (2022): *Early factory legislation*. Available at: https://www.parliament.uk/about/living-heritage/transformingsociety/livinglearning/19thcentury/overview/earlyfactorylegislation/ Accessed. 14 June 2022.

BIOGRAPHY

Patrick Zalda is currently pursuing a Doctorate at Heriot-Watt University, Scotland, UK. His research explores how ethical considerations can be embedded into the implementation of artificial intelligence and examines the broader socio-economic implications of AI in modern society. In addition, he works as a manager at a global management consultancy, where he actively applies AI governance frameworks within corporate environments.

DOI: 10.47556/B.OUTLOOK2025.23.7