Artificial Intelligence and the Ethics of Tomorrow

Brenda van Wyk and Marlene Holmner
University of Pretoria, South Africa

Brenda.vanwyk@up.ac.za
marlene.holmner@up.ac.za

Abstract: Traversing our digital information society safely and responsibly rests mainly on our comprehension of the vast sociotechnical nature of AI ethics risks, its implications and consequences. Ultimately, we all would prefer to live in a mature information society that is technologically just, inclusive and sophisticated, firmly rooted in ethical information philosophy and values. In this paper the findings of a scoping review of recent reported research look, in particular, at the sociotechnical changes and impact that disruptive AI innovation has on societies, and how this could impact new and futuristic nuances in AI ethics. The study delves into the interdisciplinarity of AI ethics. The role of intergovernmental collaboration in researching and availing frameworks and guardrails in upholding AI ethics is critically interrogated and explored. The study alludes to gaps in current research around AI ethics and impresses the need to deliberate on future AI ethics dimensions. The prerequisites for fostering further confidence and trust in AI technology are synthesised. The study concluded that inclusivity and justice in AI Ethics is not yet achieved on a global level, and that there is still a tendency towards cultural and other biases in designing, planning, implementing and also regulating AI. More research is needed on the impact and trends of AI innovation in the Global South compared to the Global North.

Keywords: AI, AI Ethics, AI frameworks, AI interdisciplinarity

1. Introduction

The sociotechnical nature of AI and GenAI remains uncontested, as AI technologies are deeply entwined in societal factors, human behaviour, and ethical considerations. Navigating the current evolving AI environment requires technical skills as well as an awareness and understanding of AI ethics, the confounds of ethical risk and transgression consequences. The recent hype around AI, conversational AI and more specifically, generative AI (GenAI) may create a false impression that AI is the ultimate panacea for the future (Floridi 2020). On a more realistic note, Floridi (2019; 2020), as well as Sartori and Theodorou (2022) remind of the many challenges as well as breakthroughs experienced over nearly sixty decades of AI existence and innovations. The remarkable AI expansion and advancement of late, particularly in GenAI, is commonly referred to as an 'AI boom' or 'AI Spring' (Floridi, 2018, Sartori & Theodorou, 2022). The phenomenal advancement in large language models (LLMs) and prompt engineering are at the core of this hype, giving raise to new ethical concerns around cultural and language biases, but also on a personal and individual level. Finding answers for ethical dilemmas for the future should start with revisiting the philosophical roots of information ethics and AI ethics, and examining current trends.

1.1 AI and the Philosophies Underpinning AI Ethics Research

The study of AI ethics is considered part of applied ethics, and on a broader level, it is a branch of the philosophy of information. Floridi (2015) defines the philosophy of information as a field of philosophy specifically looking at critical research into the conceptual nature of information and its basic principles. Moral tenets and theories of ethics pertaining to all the aspects of AI can be complex, Revisiting the discourses of seminal authors affirms the foundational ethics knowledge, that could assist in addressing the ever increasing moral and ethical issues of AI.

1.2 Corner Stones in AI Development

The origins of the ideas on information ethics, which includes AI ethics, can be traced back to the early work of scientists and mathematicians, who interrogated ideas around computer ethics. One of these questions was: can computers be held responsible for their own acts, like humans are held responsible? In 1948 Norbert Wiener offered one of the first influential models in information ethics. His work laid the foundation for further discussions on the ethical implications of information and communication technologies. It addressed issues such as control, communication, and the responsibilities associated with the use of information in automated systems.

In early 1950, the computer scientists and futurists, Alan Turing he introduced the now infamous, 'Turing Test' to assess the intelligence of machines (computers). The Turing Test consisted of a human as judge who engages in a natural language conversation with both a human and a machine, without knowing which is which. He argued that the ability to engage in natural language conversation was a reasonable criterion for attributing

intelligence to a machine. The Turing test is once again under discussion today. What seemed like science fiction in 1950 is now our lived experience with the development of LLMs used in GenAI tools and applications such as ChatGPT. For the foreseeable future, values and norms remain humanistic and not yet exclusive technological constructs. Han (2022) reports that the oldest ethics and the ethical use of information in a digital environment are computer and information ethics, which in turn paved the way for new forms of ethics and new nuances in applied ethics. A seminal workshop was organised in 1956 by an interdisciplinary team of scientists in Dartmouth in USA, where the concept of artificial intelligence was coined (Sartori & Theodorou, 2021).

It is clear that the application of moral and ethical theories to AI technology can become intricate and complex. Earlier, Bibber's (1999) approach in addressing ethical concerns centred around negative normative constructs associated with technology, looking for what is undesirable, dehumanising, holds risks, or what deepens inequalities and creates exclusion and divisions. Recent views centre around what is desirable for the greater good of all within an information society.

Prominent frameworks developed from early research to inform new approaches in ethics, information ethics, computer ethics and ethics of technology research, teachings and practices. Just some of the noteworthy examples include the OECD's Fair Information Practice and principles (FIPPs) (1980), Floridi's Information Ethics frameworks, the Mason's (1986) PAPA Framework, and Tavani's Informational Privacy Framework (1999). Here, FIPPs are regulatory guidelines for safeguarding personal data. While Floridi's framework is more philosophical, and Mason's PAPA addresses specific ethical issues. Tavani's framework balances privacy with technological benefits. Though these examples are distinctly different and contribute their own unique focuses, they all add insights into the ethical handling of information, protecting privacy, ensuring accuracy, and promoting transparency and accountability.

Floridi (2016; 2018) alludes to the erroneous use of the term information society and presenting it as a single uniform place. In reality there is a distinction of several information societies existing on different levels of maturity with salient and variant human expectations and experiences. These rapid developments necessitate a more futuristic approach in deliberating the demands of the ethics of tomorrow. As such, the rationale of this review is to gauge the extent to which recent studies are offering evidence to inform what future ethics must guard.

2. Research Design

There has been a proliferation of published research on AI and AI ethics during the past six years. The purpose of this review is to gauge the coverage of recent peer-reviewed articles and conference proceedings to get a clear indication of the disciplines and their respective focuses on the influence of AI developments on ethics. The review considers interdisciplinary studies conducted between 2018 and 2023 on aspects of AI ethics. The question to be answered is: to what extent do existing frameworks and guidelines offered by international and intergovernmental organisations inform and guide AI ethics now and in the future? This study employs a scoping review for evidenced knowledge synthesis of articles with a strong philosophical underpinning.

2.1 The Search Strategy

Peer-reviewed articles in English, published between 2018 and 2023. Answering to the following criteria: (1) the context of the discussion is an AI-based or GenAI-based, (2) the discussion includes ethical considerations and how to implement or operationalise them into the AI-based application. Based on the recent escalation AI and GenAI developments (as reported by Floridi, 2019 and others) the study is limited to the past six years.

As an inter-disciplinary review, the disciplines with AI tangents and ethics were selected. The aim of this study is to consider a more generic and macro level look at the responsible use AI ethics in information societies. The search strategy covers philosophy, ethics, information science, business studies and computer science data bases. The databases included in the search strategy in this study are Google Scholar, Scopus and Web of Science. Databases were queried during June 2024. Databases were searched using keywords and search strings representing the two main concepts of AI technology and AI ethics. Websites were searched using the local search functionality and a combination of the following keywords: "Artificial Intelligence", "ethics", "Information ethics", "the philosophy of information ethics" and "Ethics of artificial intelligence" as systematically as possible within the search functionality and capabilities of each website. These results were then filtered according to document type and then further refined to look for aspects such as interdisciplinarity, applied philosophy of ethics, intergovernmental initiatives, inclusivity and the information society. A data log of the search strategies and results with links are archived in an Excel spreadsheet.

2.2 The Selection of Evidenced Sources

Guided by the PRISMA Extension for Scoping Reviews (Tricco et al. 2018), the flow chart in Figure 1 explains the knowledge synthesis of sources.

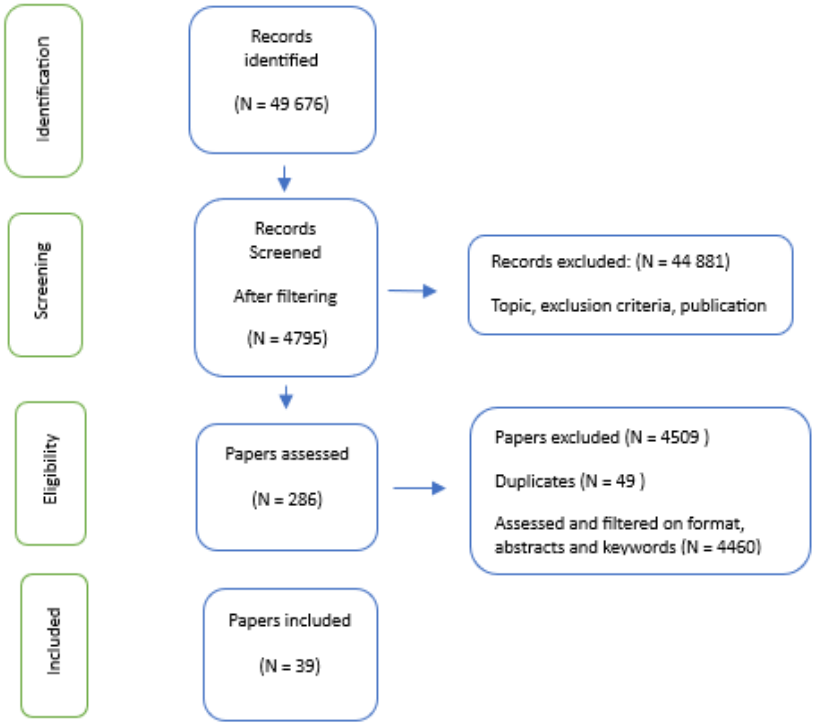

Figure 1: Scoping Review Process Flowchart

2.3 Inclusion and Exclusion Criteria

Inclusion criteria: Studies reported in full-text, peer-reviewed articles between 2018 and 2023, covering concepts of AI Ethics in the titles, abstracts or keywords were identified. Only studies in English were considered and selected.

Exclusion criteria: The following documents were excluded: grey literature, articles with abstracts only, patents, and reports, duplicate studies found in the databases.

2.4 Knowledge Synthesis of Eligible Sources

The first searches on Google Scholar, Web of Science and SCOPUS produced a result of 49 676 documents. After filtering, this was reduced to 4795 documents. In-depth screening for constructs of interdisciplinarity, philosophical and theoretical approaches, frameworks and guidelines used and reported, followed. With a total of 286 papers deemed to be eligible for this study, abstracts and keywords were screened for suitability of papers, resulting in 39 papers selected for the review. Reference links to included papers are only added to Table 1.

Table 1: Evidenced Sources included and used

Years	Types of research	Coverage	Reference
2019 -2022	Systematic literature reviews Literature reviews Typology research	A. Philosophy, Paradigms and Frameworks	A.1 Morley et al. (2020); A.2 Heiliner, (2022); A.3 Rességuier & Rodrigues (2021); A.4.Lauer, (2021); A.5 Hickok, (2021); I.2 Kazim & Koshiyama (2021);

Years	Types of research	Coverage	Reference
	Normative reflection		Also covered in: B.2 Hagendorff, 2020; C.3 Stenseke, 2022; C.4 Patel et al. (2019);
2018-2023	Systematic literature reviews Ethnographic research	B. Governance, guidelines and frameworks	B.1 Floridi (2018); B.2 Hagendorff, 2020; B.3. Butterworth, (2018) Also covered in: A.2 Morley et al. 2020; C.1 Bisconti et al (2023); F.2 Wright (2023)
2019 - 2023	Collaborative research	C. Interdisciplinarity, multidisciplinary	C.1 Bisconti et al (2023) ; C.2 Stenseke, 2022; C.3 Patel et al. (2019) ; C.4 Reeve, Zaugg & Zheng (2022)
2022 - 2023	Literature review	D. Education	D.1 Corrigan et al (2023); D.2 Raji, Scheuerman & Amironsei(2021) ; D.3 Bannister, Santamaría-Urbieta & Alcalde-Peñalver (2023); D.4 Holmes et al (2022); Also covered in C.5 Reeve, Zaugg & Zheng (2022); H.2 Chan (2023)
2019-2023	Systematic literature review	E. Sociotechnical nature of AI ethics	E.1 Sartori & Theodorou, (2022); E. 2 Floridi (2019); E.3 Green, 2021; E.4 Hagerty & Rubinov, (2019). Also covered in: G.1 Corrigan et al (2023); C.4 Patel et al. (2019);
2018 - 2013	Content analysis	F. Information Society, Mature information societies	F.1 Floridi et al (2018); F.2 Wright (2023); F.3 Bruneault & Sabourin (2021);F.4 Floridi (2020).
2023	Literature review Narrative inquiry	G. Justice and Inclusivity	G.1 Roche, Wall & Lewis, (2023); G.2 Hoffmann (2021); G.4 Park & Humphry (2019) Also covered in: D.2 Raji, Scheuerman & Amironsei, (2023); D.3 Bannister, Santamaría-Urbieta & Alcalde-Peñalver (2023);
2023	Bibliometric analysis	H. Generative AI Ethics	H.1 Bozkurt (2023) H.2 Chan (2023); H.3 Stahl & Eke (2023).
2021-2023	Systematic literature reviews	I. Human control, Human in the loop/ society in the loop	I. 1 Tasioulas; 2022; I.2 Kazim & Koshiyama; I.3 Akintunde (2023); I.4 Kyriakou & Otterbacher (2023); I.5 Riikka(2021); 1.6 Heyder et al (2023) Also covered in: C.1 Bisconti et al (2023)
2021	Bibliometrics	J. Digital identity	J.1 Zhang et al (2021); J.2 Cioroaica (2022), J.3 Bedushi (2019)

3. Discussion of Findings

The types of research studies within the final results included systematic literature reviews, scoping reviews, autoethnographic research, narrative inquiries, case studies, bibliometric studies and crowdsources research. Here, crowdsourced research is a model of the social organisation of research as a large-scale collaboration in which one or more research projects are conducted by multiple teams in an independent, yet coordinated manner. There has been a sharp increase in the number published studies on the topic of AI ethics in the Scopus database from 2021 to date.

3.1 Representative Countries

The results of this review indicated that most studies that met the criteria for inclusion, were conducted in countries within the Global North. These include the United Kingdom, USA, Canada, and Australia. This is in line with the extensive bibliometric study by Zhang et al (2021). They add that comparable research is being conducted in China and the European countries, but their key research communication is less accessible due to language barriers. There is a notable paucity in reported research emanating from Global South countries.

3.2 A. Philosophy, Paradigms and Frameworks in Researching AI Ethics

Many of the selected studies do not mention or elaborate on a particular paradigm or framework that informed their research. In their study, Roche et al (2021) profess an intersectional analytical framework to address the inherent complexities of AI ethics, and state that it is possible to understand how aspects of social and political identities combine to create different modes of discrimination and privilege.

3.3 B. Governance, Guidelines and Frameworks

Intergovernmental organisations, such as the OECD, UN and UNESCO offer valuable insights and guardrail on dimensions of AI ethics such as the inclusivity, education, awareness, towards further cooperation among countries and regions. In this review research (including Roche et al., 2022) report on the continuous efforts to address AI ethics risks, more particularly on social and cultural levels, resulting an ever-growing body of new knowledge that must translate into AI policies, frameworks and guidelines.

As early as 1980 the OECD's Privacy Guidelines outlined principles for the protection of privacy and personal data. Yet, today authors are critical and questions the intentions and contribution of intergovernmental organisations and their initiatives (Roche et al, 2021; Green, 2021; Floridi, 2018; Wright, 2023).

Wright (2023) is sceptical of organisations, governments and intergovernmental organisations who blatantly make use of “ethics-washing” practices in an effort to delay regulation of the sector. Green (2021) is in agreement, and warns of the gap between ethics as a mode of normative inquiry and ethics as a practical endeavour. There are suggestions of hiding political and normative disagreements. Rességuier and Rodrigues (2021) calls these guidelines “toothless” and ineffective. Conversely, Floridi (2018) argues that baseless and misplaced concerns may lead to underused AI technologies.

On the plus side, member countries have enacted legislation based on the information and guardrails offered, intended to protect individuals against the risk of AI abuse, and still give individuals the right of access to data with a view to checking their accuracy and appropriateness (Wright, 2023, 2021). The 2013 Privacy framework by OECD, and later in 2023 the OECD's Good Practice Principles for Data Ethics in the Public Sector serve as examples.

Hagerty and Rubinov (2019) underpin the need to ensure that emerging frameworks and standards themselves do not reproduce and reinforce a variety of biases, nor augment inequality. Authors (Hagerty & Rubinov, 2019; and Roche et al , 2021) allude to a lack of diversity as it is predominantly the research from Global North countries and research institutions are spearheading new developments in AI technologies and algorithms, bringing their own socio-cultural lenses. This homogenisation of applications in itself add to further exclusions and ethical biases.

3.4 C. The Interdisciplinarity, Multidisciplinary and Intersectionality of AI and GenAI Ethics

Patel et al (2019) state that by design, AI ethics must be interdisciplinary, as it considers broad pathways covering different disciplines to navigate AI innovation and ethics. Sartori and Theodorou (2022) add intersectionality as an important construct to the AI ethics debate. They contest that people from different backgrounds experience the world differently, be it due to gender, class, race or other forms of identity. This in turn affirms Floridi’s argument that we are dealing with a multiple of information societies. Stenseke (2022)

refers to these challenges as 'grand challenges' of global importance, impacting all aspects of human life, making the interdisciplinary approach critical.

3.5 D. Education

The centrality of education and training to the AI ethics debate delivered results within the search. Holmes et al (2022) remind that the core aim of education is to support learning with technology assisted teaching. AI concerns include fairness, accountability, transparency, bias, autonomy, agency, and inclusion (Holmes et al 2022). The compounded ethical challenges are present in the possibility of unintended consequences. It is important to distinguishing between differentiate between doing ethical things and doing things ethically in pedagogical choices.

3.6 E. The Sociotechnical Nature of AI and GenAI Ethics

Affirming the earlier studies of Bibber (1999) and Moore (2006), studies continuously report on the sociotechnical nature of AI that cannot be negated. The humanistic approach lies mainly in the sharing of values. Roche et al (2021) add that the rapid advancement of AI could have a positive and a transformative impact in individuals circumstances, as well as advancements in the broader economy.

3.7 F. Using AI Ethics Towards Maturing Information Societies

Florida (2019) criticises intergovernmental organisations' over-emphasis on technology. He laments that their narrow interest and remit focus on areas of development in broadband and telecommunications, the internet, consumer policy and digital governance. Floridi (2018) stresses that a mature information society centres around expectations. He states there is no global universal experience around digital innovation in the information society. We need to shift from viewing activities as 'off line' or 'online', to where we experience our world 'onlife' in an environment that Floridi (2018) calls the infosphere.

3.8 G. Justice and Inclusivity

The central ethics themes of AI include privacy and transparency. Hickock (2021) reports perspectives from different regions, and communities vary and the view that there are global and universal approaches to ethics is incorrect. Roche et al (2022) warn of ethnocentricity, whereby other cultures are evaluated through the standards and perspectives of the dominant culture. The Global South has to deal with historically unequal, colonial and biased social and power structures.

Reeve, Zaugg and Zheng (2022) warn that real-world harm is rife with biases such as racial profiling in online systems. Linking to education and interdisciplinarity of AI ethics teachings must urgently happen in computer science, data science, statistics, the social sciences and the humanities. Hagerty and Rubinov (2019) state that AI can deepen social divides, and exacerbating social inequality, particularly among historically-marginalized groups.

3.9 H. GenAI, Predictive AI and Conversational AI

Both GenAI and conversation AI , though distinctly different, are associated with AI risks. With rapid developments in GenAI, like ChatGPT, the need for human skills training is at the forefront. With prompt engineering attempting to make sense of human prompts, the concern for ethical problems increases. Immediate concerns are the risk to access and fairness. The AI and ethics of AI literacy communities should also focus on training individuals on the cognitive skills needed to become optimal users of such technologies. Bozkurt (2023) hails the educational potential of AI, but warns that it should be approached with caution to ensure responsible and ethical use.

3.10 I. Positioning Human Control in AI Autonomous Capabilities

The autonomous nature of AI is a huge concern for many, and there is a prevailing apprehension and fear for exclusion of human control. Kyriakou and Otterbacher (2023) emphasise the critical synthesis around the matter of human oversight. They refer to Rahwan who coined the human-in-the-loop (HITL) concept, and later society-in- the- loop (SITL) where human oversight can regulate AI and algorithm application and design as a form of social contract. Riikka (2021) alludes to inherent limitations and human error, and the inability to provide protection for fundamental rights. Akintunde (2019) states that replacing human decision-making with machine decision-making will be met with distrust in AI systems and moves to keep the human user in the loop.

3.11 J. Digital Identity, Ownership, Consent and Control

Beduschi (2019) warns that digital identity platforms may infringe human rights, and may infringe on privacy. It has the potential to be used in a discriminating manner. As digital identity refers to data informing our networked identity that uniquely describes and contains detailed information about a person, entity or business. Our digital footprint necessitates risk mitigation on both sociological as well as psychological levels. The particular AI threats of identity theft, deep fakes, to name but a few . Autonomy lies in user consent, ownership and control of data. Cioroaica et al (2022) posit that the need for ethical digital identities to fill the trust gap.

4. Conclusion

This review did a macro level exploration (see as explained Woodgate & Ajmeri, 2024) on the responsible use AI ethics, and also explore what is in stall for the near future. The development of AI has escalated significantly over the past six years. Research by scientists like Floridi and Satori, shows how deeply sociotechnical realities are interwoven in digital societies, and impress on us the need to think futuristically.

There are not many studies considering future AI ethics. Studies still concentrate on dealing with existing ethics challenges. The effectiveness of AI ethics to mitigate future harm lies in global efforts to sustain mature and informed information societies. Roche et al (2021) posit that AI ethics policies and standards must align with a variety of socio-cultural and socio-economic context. Regulation and accountability of AI ethics must continue to be visible and transparent.

The need for informed and clear AI ethics guidelines are bound to become increasingly complex, requiring foresight and expert knowledge of the technology, psychological and societal factors. AI as an emerging technology, has presented with various uncertainties, and the moral reaction on these developments are in many instances still fully or partially unknown. As we progress deeper into this era of technology, it is crucial to maintain a balance between innovation and accountability.

5. Recommendations

Based on the findings of the review, recommendations are made on how best to prepare to ensure the responsible use of AI in the future. The study recommends further research on how foresight analysis can be deployed in AI technology developments.

Floridi (2019) posits that predicting the ethics for the foreseeable future is often done with hindsight knowledge. There is a trend to build on fear and uncertainty:

> *"On the untestable side, some people will keep selling catastrophic forecasts, with dystopian scenarios taking place in some future that is sufficiently distant to ensure that the Jeremiahs will not be around to be proven wrong" (Floridi, 2019, 1)*

In a developing scenario, what the full extent of this maturity is, is not yet known. The infiltration of political and economic unethical opportunism in activities, finding their supple way into ethics frameworks and guidelines must have consequences. There are many unknowns, and Floridi (2019) states that it will not be innovation and change per se, but the regulation thereof that will determine the ethics of AI in the near future.

As AI advancements involve increased autonomous AI decisions, the singularity discourse must be on the radar of future ethics planning and frameworks. Floridi and Strait (2020) propose that ethics foresight analysis become more widely used. Ethical Foresight Analysis (EFA) is not a new concept in research and development, it refers to a variety of analytical methodologies for anticipating or predicting the ethical issues that technological artefacts may raise (Floridi & Strait, 2020).

References

Akintunde, M., Young, V., Yazdanpanah, V., Fathabadi, A.S., et al (2023). Verifiably Safe and Trusted Human-AI Systems: A Socio-technical Perspective. In Proceedings of the First International Symposium on Trustworthy Autonomous Systems (TAS '23). Association for Computing Machinery, New York, NY, USA, Article 56, 1–6. https://doi.org/10.1145/3597512.3599719

Bannister, P., Santamaría-Urbieta, A., & Alcalde-Peñalver, E. (2023). A Delphi Study on Generative Artificial Intelligence and English Medium Instruction Assessment: Implications for Social Justice. *Iranian Journal of Language Teaching Research*, *11*(3 (Special Issue)), 53-80. Doi: 10.30466/ijltr.2023.121406

Bindé, J., (2001) Toward an Ethics of the Future". *Globalization*, edited by Arjun Appadurai, New York, USA: Duke University Press, 2001, pp. 90-113. https://doi.org/10.1515/9780822383215-007

Beduschi, A. (2019). Digital identity: Contemporary challenges for data protection, privacy and non-discrimination rights. *Big Data & Society*, *6*(2). https://doi.org/10.1177/2053951719855091

Butterworth, M., (2018). The ICO and artificial intelligence: The role of fairness in the GDPR framework,

Computer Law & Security Review, 34 (2): 257-268, https://doi.org/10.1016/j.clsr.2018.01.004

Bisconti, P., Orsitto, D., Fedorczyk, F. *et al.* Maximizing team synergy in AI-related interdisciplinary groups: an interdisciplinary-by-design iterative methodology. *AI & Society,* **38**, 1443–1452. https://doi.org/10.1007/s00146-022-01518-8

Bozkurt, A. (2023). Unleashing the Potential of Generative AI, Conversational Agents and Chatbots in Educational Praxis: A Systematic Review and Bibliometric Analysis of GenAI in Education. Open Praxis, 15(4), pp. 261–270. DOI: https://doi.org/10.55982/openpraxis.15.4.609

Chan, C.K.Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *Int J Education Technology Higher Education* **20**, 38. https://doi.org/10.1186/s41239-023-00408-3

Cioroaica, E., Buhnova, B., Jacobi, F and Schneider,D., (2022). The Concept of Ethical Digital Identities, *IEEE/ACM 1st International Workshop on Software Engineering for Responsible Artificial Intelligence (SE4RAI)*, Pittsburgh, PA, USA, 2022, pp. 17-20, DOI: 10.1145/3526073.3527586.

Corrigan, C.C et al. (Ed.) (2023) *AI Ethics in Higher Education: Insights from Africa and Beyond*, Springer International Publishing AG, 2023.

Floridi, L. (2013). *The philosophy of information*. Oxford University Press.

Floridi, L. (2015). *The ethics of information*. Oxford University Press.

Floridi, L. (2016). Mature Information Societies—a Matter of Expectations. *Philos. Technol.* **29**, 1–4 (2016). https://doi.org/10.1007/s13347-016-0214-6

Floridi, L. (2018). Soft Ethics and the Governance of the Digital. Philosophy & Technology, 31(1), 1–8. https://doi.org/10.1007/ s13347-018-0303-9

Floridi, L., Cowls, J., Beltrametti, M. *et al.* (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines* **28**, 689–707. https://doi.org/10.1007/s11023-018-9482-5

Floridi, L. (2019). What the Near Future of Artificial Intelligence Could Be. Philosophy & Technology, 32(1), 1–15. https://doi. org/10.1007/s13347-019-00345-y

Floridi, L. (2020) AI and Its New Winter: from Myths to Realities. *Philos. Technol.* **33**, 1–3. https://doi.org/10.1007/s13347-020-00396-6

Floridi, L., and Strait, A. (2020). Ethical Foresight Analysis: What it is and Why it is Needed? *Minds & Machines* **30**: 77–97. https://doi.org/10.1007/s11023-020-09521-y

Green, B (2021). The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics in Practice, in *Journal of Social Computing*. 2(3): 209-225, September 2021, Doi: 10.23919/JSC.2021.0018

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines* **30**, 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8

Han, J. (2022). An Information Ethics Framework Based on ICT Platforms. *Information*, *13*, 440. https://doi.org/10.3390/info13090440

Heilinger, J.C. (2022). The Ethics of AI Ethics. A Constructive Critique. *Philos. Technol.* **35**, 61. https://doi.org/10.1007/s13347-022-00557-9

Hickok, M. (2021). Lessons learned from AI ethics principles for future actions. *AI Ethics* **1**, 41–47. https://doi.org/10.1007/s43681-020-00008-1

Hoffmann, A. L. (2021). Terms of inclusion: Data, discourse, violence. *New Media & Society*, *23*(12), 3539-3556. https://doi.org/10.1177/1461444820958725

Holmes, W., Porayska-Pomsta, K., Holstein, K. *et al.* (2022). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal Artifificial Intelligence Education,* **32**, 504–526. https://doi.org/10.1007/s40593-021-00239-1

Kazim, E and Koshiyama, A.S, (2021). A high-level overview of AI ethics, Patterns, 2 (9), https://doi.org/10.1016/j.patter.2021.100314

Lauer, D. (2021). You cannot have AI ethics without ethics. *AI Ethics* **1**, 21–25. https://doi.org/10.1007/s43681-020-00013-4

Mason, R. O. (1986). Four Ethical Issues of the Information Age. *MIS Quarterly*, *10*(1), 5–12. https://doi.org/10.2307/248873

Morley, J., Floridi, L., Kinsey, L. *et al.* (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science Engineering Ethics* **26**, 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

OECD, (2013) The OECD Privacy Framework, https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf

Park, S., and Humphry, J. (2019). Exclusion by design: intersections of social, digital and data exclusion. *Information, Communication & Society*, *22*(7), 934–953. https://doi.org/10.1080/1369118X.2019.1606266

Patel, M., Webb., H, Jirotka, M., Davoust, A., Gales, R., Rovatsos, M and Koene, A. (2019). Harnessing Interdisciplinarity to Promote the Ethical Design of AI Systems. in P Griffiths & M Nowshade Kabir (eds),ECIAIR 2019 - Proceedings of European Conference on the Impact of Artificial Intelligence and Robotics. Academic Conferences and Publishing International (acpi), European Conference on the Impact of Artificial Intelligence and Robotics, Oxford, United Kingdom, 31/10/19.

Raji, I.D, Scheuermann, M.K., and Amironesei, R. (2021). You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education, FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, March 2021: 515 – 525, https://doi.org/10.1145/3442188.3445914

Reeve, J., Zaugg, I. and Zheng, T. (2022), "Mapping data ethics curricula", *Journal of Information, Communication and Ethics in Society*, 20 (3): 388-399. https://doi.org/10.1108/JICES-12-2021-0124

Sartori, L., Theodorou, A. (2022). A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. *Ethics Information Technology,* **24**, 4 (2022). https://doi.org/10.1007/s10676-022-09624-3

Stahl, B.B. and Damian Eke, D. (2024). The ethics of ChatGPT – Exploring the ethical issues of an emerging technology, *International Journal of Information Management*, Volume 74, https://doi.org/10.1016/j.ijinfomgt.2023.102700

Stenseke, J. (2022). Interdisciplinary Confusion and Resolution in the Context of Moral Machines. *Science Engineering Ethics* 28, 24. https://doi.org/10.1007/s11948-022-00378-1

Tasioulas, J. (2022). Artificial Intelligence, Humanistic Ethics. *Daedalus* 2022; 151 (2): 232–243. Doi: https://doi.org/10.1162/daed_a_01912

Tavani, H.T. (1999). Informational privacy, data mining, and the internet. *Ethics and Information Technology,* **1**(2), 137–145.

Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. (2018) PRISMA Extension for Scoping Reviews (PRISMAScR): Checklist and Explanation. Ann Intern Med. 2018;169:467–473. Doi: 10.7326/M18-0850

Woodgate, J., and Ajmeri, N. (2024). Macro Ethics Principles for Responsible AI Systems: Taxonomy and Directions. ACM Computing Surveys, 1-39. Advance online publication

Wright, J. (2023). The Development of AI Ethics in Japan: Ethics-washing Society 5.0? *East Asian Science, Technology and Society: An International Journal*, 1–18. https://doi.org/10.1080/18752160.2023.2275987

Zhang, Y., Wu, M., Tian, G.Y., Zhang, G and Jie Lu, J, (2021). Ethics and privacy of artificial intelligence: Understandings from bibliometrics, *Knowledge-Based Systems*, 222: 2021, https://doi.org/10.1016/j.knosys.2021.106994