Chapter 4

Artificial Intelligence and the Ethics of Tomorrow: Tracing the Shift in Information Ethics through History

Brenda van Wyk
Department of Information Science
University of Pretoria
Pretoria, South Africa

Marlene Holmner
Department of Information Science
University of Pretoria
Pretoria, South Africa

Introduction

The emergence and ongoing development of the information era have given rise to various ethical dilemmas that are inherent to the essence of information. IE (information ethics) is an interdisciplinary area that incorporates disciplines such as philosophy, computer science, sociology, law, and others, as highlighted by scholars like Quin (2011) and Floridi (2019). Although the information age is frequently regarded as a contemporary occurrence, this perspective fails to acknowledge its underlying historical origins. The origins of this age can be attributed to previous advancements in communication and Its (information technologies). The information age is currently recognised as a dynamic and ongoing era characterised by swift technological developments that are continuously transforming various aspects of society, such as communication, education, healthcare, and entertainment.

IE is a discipline that combines ethical theories with practical considerations about technology. It acknowledges the intricate and diverse character of ethical dilemmas in the digital world. The subject matter encompasses inquiries on data privacy, cybersecurity, digital rights, and the ethical utilisation of AI (artificial intelligence) and ML (machine learning). With the advancement and widespread use of technology, ethical questions become more complex and important.

Furthermore, IE is dynamic, constantly adjusting to address novel issues and inquiries presented by evolving technologies. For example, the emergence of social media platforms has ignited issues on misinformation, digital identity, and online conduct, while progress in AI and robotics is stimulating conversations about machine ethics, job displacement, and the prospects of human-machine interactions in the future.

As we traverse this changing environment, it is crucial to comprehend and tackle the ethical consequences of our digital existence. In this chapter the findings of a review of recent literature allude to the need to address issues of the privacy and security issues related to data collection, digital identity and authenticity, accessibility and digital inclusivity, legal frameworks, surveillance, and the freedom of expression. The study delves into how AI innovation influences new nuances in IE. The ultimate goal is to establish a world that is not only technologically sophisticated but also firmly rooted in ethical values.

Contribution of Information Science and Informatics to Information

Ethics

IE draws upon diverse disciplinary insights and historical views to navigate the current revolutionary era. Its aim is to ensure that technical advancements are in harmony with our shared values and societal objectives. Within this environment, it is of utmost importance to thoroughly examine and analyse ethical concepts and frameworks. This examination plays a vital role

in determining the course of action for creating a future where technology is utilised to benefit humanity's highest ideals.

With its early beginnings in 1937, ASIS (the American Society for Information Science – later renamed to the American Documentation Institute) explored aspects of information and information behaviour (Alharbi & Mukhari 2023:38). Two decades later Borko defined information science as the discipline exploring what information and information seeking behaviour are, as well as the flow of information as it is processed for accessibility and use (Borko 1968). However, the person who coined the term *information ethics* was Robert Hauptman (Froehlich 2000:264).

Information science is concerned with that body of knowledge relating to the origination, collection, organisation, storage, retrieval, interpretation, transmission, transformation, and utilisation of information. Theorists agree on the interdisciplinarity of information science (Wilson 1997; Stahl 2008; Bates 2005). It has both a pure science component, which enquires into the subject without regard to its application, and an applied science component, which develops services and product.

What is Information Ethics?

IE, in it its simplest form, refers to the use of information in a way that it does not cause harm. As technology evolves, new ethical challenges and dilemmas arise (Moor 1985; 2020; Burgess & Knox 2019). As a subset of information science and being part of applied ethics, IE examines ethical issues associated with the creation, dissemination, and use of information. It is particularly relevant to the informatics discipline that deals with the management and analysis of information.

On a macro level Zhou, Chen, Berry, Reed, Zhang, and Savage (2020:3010) share that ethics as a branch of philosophy refers to the systematising, defending, and recommending concepts of right and wrong conduct, where rights, obligations, benefits, and fairness principles are explained and deliberated in models and frameworks.

The foundations of IE are instituted in its key principles and constructs. Figure 4.1 offers a summary of the overlapping constructs included in theories and research that will be discussed in greater detail this chapter.

Figure 4.1: Summary of IE constructs. (Source: Personal archive)

These foundational principles guide the development of ethical frameworks, policies, and practices in the rapidly evolving landscape of IT and data management**.** Adhering to these principles helps to ensure that information is used responsibly, ethically, and in ways that benefit individuals and society. IE is at the heart of AI and the ethics of AI (Bester & Fisher 2020:2). These ethics are underpinned by the philosophy of information.

In the Beginning: The Philosophy of Information and Information Ethics

PI (the philosophy of information) encompasses aspects of philosophy, computer science, information science, cognitive science, and communication studies. It addresses fundamental questions related to the nature of information, its role in the universe, and its impact on knowledge, reality, and society (Bynum 2010:420). As an interdisciplinary field, it explores the

nature, properties, and implications of information. Floridi (1999:41) defines PI as a field of philosophy, more specifically looking at critical research into the basic principles of information and the conceptual nature thereof. This field includes the dynamics of information, its uses and applications theory, and computational methodologies (Floridi 2013:38; 2015:42). IE, in turn, is a branch of PI. New developments, such as AI and Gen-AI (generative artificial intelligence), have a profound impact on moral decision-making in our daily lives.

Moral tenets and theories can be complex, and when revisiting the discourses of seminal authors it lays a foundation to meaningfully expand on the ever increasing moral and ethical issues of AI. Floridi (2015:197) stresses the importance to find a balance in practices and experiences in daily lives and uses the metaphor of a tree depending on its roots for the sustainable growth of new branches. Seminal theorists of IE are particularly found in mathematics (Shannon 1948), informatics (Turing 1950), and information science (cf. Mason 1986; Capurro, Eldred, & Nagel 2013).

Finding the Roots of Information Ethics

On the challenges encountered when applying human-centred philosophy of morals and ethics in ICTs (information and communication technologies), Birrer (1999:16) laments: 'It is by no means accidental that most of the discussion about values in technology focuses on normative components that are considered undesirable: Technocracy, dehumanization, inequality, manipulation, loss of control.' Values and norms are humanistic and not technological constructs (Moor 2006:18). As such, applying moral and ethical theories to information and ICTs is complex.

Expanding upon the first ideas about the essence and intricacies of AI, we arrived at the overarching question: *Can computers, like humans be held responsible for their acts?* One of the notable futurists and mathematicians was Alan Turing. In the early fifties of the previous century he was a pioneering mathematician and computer scientist, who introduced the *Turing*

Test to assess the intelligence of machines. In 1950 his book, *Computing machinery and intelligence* (Turing 1950) posed and addressed the questions on whether computers can 'think.' The Turing Test involves a human judge engaging in natural language conversations with both a human and a machine, without knowing which is which. If the judge cannot reliably distinguish between the human and the machine based on the conversation, then, according to Turing, we could say that the machine is demonstrating a form of intelligence indistinguishable from that of a human.

The Turing Test remains a widely discussed concept in AI and philosophy, and it has sparked debates about the nature of consciousness, intelligence, and the potential capabilities of machines. The Turing Test is once again in the limelight, this time on the development of LLMs (large language models) used in Gen-AI.

Early Philosophers and Pioneers of the Information Ethics Theory

One of the first and influential models in IE is associated with the work of Norbert Wiener. In 1948, Wiener, a mathematician and philosopher pioneered research in cybernetics, a precursor to IE. His work laid the foundation for discussions on the ethical implications of ICT. While Wiener's work may not fit the contemporary definition of a comprehensive IE model, it has set the stage for ethical reflections on the impact of technology on individuals and society, addressing issues such as control, communication, and the responsibilities associated with the use of information in automated systems.

IE Models and Frameworks

Prominent frameworks that have been informing IE research, teachings, and practices include, among others, the 2013 privacy framework by the OECD (Organisation for Economic Cooperation and Development) FIP (fair information practice) (OECD 2013), Floridi's IE frameworks, the PAPA (privacy, accuracy, property, and accessibility) framework, and Tavani's

informational privacy framework. This chapter provides an overview of IE models and frameworks and discusses whether the capability of existing frameworks is comprehensive enough to provide the ethical guardrails offered to accommodate the ethical considerations prevalent in a fast-changing world. IE typically considers examining the ethical implications and considerations associated with disruptive innovations around the creation, use, and management of information within virtual and immersive environments. The question now is to what extent exiting frameworks are still relevant to underpin the vast new developments.

While the frameworks mentioned above, provide a useful foundation for evaluating ethical issues in technology, including aspects of Gen-AI, there may be potential shortcomings requiring further adaptation to address the unique challenges posed by new developments. Han (2022:1 of 11) reports that the oldest and most referred to terms applied to the ethical use of information in a digital environment are 'computer' and 'IE.' The finding of this study is that new aspects must be considered, such as cross-cultural aspects on a global scale, devising guidelines for content moderation, ensuring that AI systems are ethical and avoid discriminatory practices, and lastly, considering environmental impact ethics. The conclusion is that one should take the wisdom of existing frames and build a comprehensive, adaptable multidisciplinary framework on it to address ethical dilemmas presented by evolving technological advancements and changing societal norms. Such a framework should encourage responsible innovation by providing the ethical guardrails, and simultaneously safeguarding the rights, wellbeing, and dignity of individuals in virtual spaces.

Floridi's Information Ethics Frameworks

Luciano Floridi is a contemporary philosopher of information, served as UNESCO (United Nations Educational, Scientific, and Cultural Organisation) chair in information and computer ethics. His framework considers the ethical implications of living in an information society. He coined the term 'infosphere,' which can be found in the context of the biosphere, which explains the

entire informational environment constituted by all informational entities, interactions, processes, and mutual relations (Floridi 2013:132).

Floridi (2016:3; 2018:2) argues that the current information society can be described as a 'mature information society,' arguing that it developed through the first- and second-order technologies and ICTs. The term 'information society' refers to a society where the creation, distribution, and manipulation of information play a significant and pervasive role in economic, social, cultural, and political activities (Floridi 2018:4). He emphasises the ethical importance of treating information ethically and respecting the rights and interests of individuals and communities in the information age. He states that the rapid development in ICTs created a new information environment, requiring a suitable ethics framework to address unprecedented challenges in the environment, where a new understanding must be created between artificial and real environments (Floridi 2010:219).

Floridi's continuing research into information creation and its ethical use with AI and AI systems positions information science and IE ethics in AI ethics research towards comprehensive models and theories in a changing field of study (Floridi 2018; 2019).

Mason and the PAPA Model

In 1986, Richard Mason published a social framework for addressing the major ethical issues of the information age in his pivotal article, *Four ethical issues of the information age* (Mason 1986:9). This framework consists of four broad categories of ethical issues namely privacy, accuracy, property, and accessibility, hence PAPA. This PAPA framework is still relatively germane in studying the ethical issues in IT (Woodward, Imboden, & Martin 2011:64). Due to the increasing prevalence of digital data and the inherent dangers connected with its storage and transfer, security becomes a crucial ethical concern in the information age. It is the responsibility of information security to protect data

integrity, confidentiality, and accessibility, immediately affecting ethical concepts like privacy, trust, and accountability.

When considering the possible harm caused by data breaches, such as identity theft, financial loss, or privacy invasion, the ethical implications of security become clear. These violations can damage the public's trust and violate people's rights (Solove 2006:478). Therefore, it is the responsibility of information professionals to safeguard the integrity of the digital ecosystem, protect stakeholders, and follow ethical standards in their security practices. It has thus become vital to include a 'S' to the PAPA acronym.

Table 4.1: The PAPAS model (Adapted from Mason 1986; Young, Smith, & Zheng 2020)

Privacy	**Accuracy**
• The expansion of ICTs has resulted in the collection and use of personal information on a huge scale. • This raises ethical considerations about who has access to and uses this information, as well as concerns about the possibility of surveillance and manipulation. • Categories of private information: Private communication, privacy of the body, personal information, and information about one's possessions.	• Data integrity becomes increasingly important, as massive databases grow more interconnected. • Technology have made it easy for false or misleading information to spread quickly and widely, which can have serious consequences for individuals and society. • How to combat misinformation and disinformation. • How to ensure accurate and reliable access.

Property · The internet and digital technologies have made it easy to share and distribute information, including copyrighted material. · The free flow of information is a risk to intellectual property and rights of creators and owners of this material. · IP (intellectual property) refers to the ownership of ideas and creative works, including patents, trademarks, and copyright. · Each of these instruments of protection is governed by a set of laws and regulations. · Some of these rules are not universal and they only apply in specific countries.	**Accessibility** · ICTs increase access to information, but equality in access is not yet achieved and not everybody can equally participate in society. · Expanded literacy and reasoning skills are essential for one's participation in the growth of any society (Mason 1986). · Access to the essential technologies is required, while information must be available to be used and consumed.
Security · Security involves such a broad variety of concerns in the context of the information age – from data breaches to cyberattacks. Information security in its most basic form aims to protect the availability, confidentiality, and integrity of data.	

AI and other emerging technologies present new vulnerabilities and the possibility of abuse, greatly complicating the security picture.

Tavani's Informational Privacy Framework

Richard Tavani, a scholar known for his work in computer ethics and IE, wrote a book, *Ethics and technology: Controversies, questions, and strategies for ethical computing* which is widely used in the field of computers and IE. Tavani has explored ethical issues related to IT, including privacy concerns. He has identified three types of privacy:

· Mental or psychological privacy;

- physical privacy, also referring to access privacy; and
- decisional and informational privacy (Tavani 1999:138-140).

In the context of informational privacy, the concept of informed consent is crucial. Users should be aware of how their information is being collected, processed, and shared, and they should have the ability to make informed decisions about whether to share their information. Tavani emphasises the principle of data minimisation, which suggests that organisations should only collect and retain the minimum amount of information necessary for a specific purpose. Frameworks addressing informational privacy often highlight the importance of giving individuals control over their personal information. This includes mechanisms for individuals to access, correct, and delete their data.

Transparency in data practices and accountability for how organisations handle information are key elements. Tavani discusses how organisations should be transparent about their data practices and accountability for any misuse of information (Tavani 1999:137). Tavani's work on privacy adds value to the study of the ethics of AI, as privacy on many levels is still a concern in the application of AI algorithms.

Charles Ess

Ess is a philosopher and scholar who has made significant contributions to the field of media ethics and IE, particularly in the context of computer and internet ethics. He studied the ethical challenges stemming from privacy, identity, digital communities, and social media platforms. He is particularly concerned with feminist ethics where gender-related ethical dilemmas such as harassment and bullying manifest in gender roles in cyberspace.

Ess advocates for the recognition of diverse cultural values in a global context. He has contributed to the theoretical foundations of IE (Ess 2008). His contributions extend to the development of global IE frameworks, emphasising the need for a global dialogue on ethical issues related to ICTs, considering diverse cultural, social, and political contexts (Ess 2014).

Ess has explored ethical implications associated with emerging technologies, including AI and robotics. His work addresses issues such as accountability, transparency, and the societal impact of intelligent systems.

Raphael Capurro

Capurro is a philosopher and information scientist who has made significant contributions to the field of IE. In 1999 he established the ICIE (International Center for Information Ethics), and currently the centre has a diverse membership covering the fields of informatics, information science, computer science, and more (Froehlich 2000). Capurro has contributed theoretical insights and frameworks that have influenced digital ontologies. His frameworks focus on freedom, privacy, and identity in an online and cyberworld (Capurro *et al*. 2013:12). Capurro warns that the tendency to see everything within the context of a Western philosophy, must be challenged and different cultural notions must be considered.

One notable contribution by Capurro is the concept of 'hermeneutical ethics of information.' Hermeneutics, in a broad sense, refers to the interpretation of meaning. Capurro emphasises the importance of understanding information within a cultural and historical context, considering the multiple layers of interpretation that can be applied to information. This perspective recognises the ethical implications of interpretation, representation, and communication in the information domain.

Understandably, the field of IE is multidisciplinary, resulting in various frameworks and models. As the scholars discussed above are not deplete, many others also contributed to the understanding of ethical issues in IT and communication. The frameworks developed over several decades and paved the way for new frameworks and discourse on the ethics of AI, focussing on the responsible use of AI and GenAI.

The Ethics of AI

The recent hype around AI and Gen-AI created the illusion that AI is the ultimate panacea (Floridi 2020:2). Views are ranging from

ultimate doom to a hailing of AI as the beginning of everything. This is not realistic and authors such as Floridi (2019; 2020) as well as Sartori and Theodorou (2022:10 of 11) remind us that during the many decades of the becoming of AI, there were many wins and losses. What cannot be denied, is the rapid tempo of recent developments of LLMs and prompt engineering. The sociotechnical nature of AI and Gen-AI remains uncontested, where the development, deployment, and impact of AI are deeply intertwined with societal factors, human behaviour, and ethical considerations.

The Nature of AI and Gen-AI

AI and AI ethics are not new fields of research and study. Floridi (2020:3), as well as Sartori and Theodorou (2022:2 of 11) point to the many 'winters' and 'summers' that AI has already endured. In 1956, a group of mathematicians and computer scientists met at the Dartmouth College, New Hampshire, to discuss the emerging field of AI (Strickland 2021). It was here that John McCarthy coined the term 'artificial intelligence,' which he argued, 'would explore the hypothesis that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it' (Strickland 2021:27). Bartneck, Lütge, Wagner, and Welsh (2020:18) allude that a complete definition of AI is still in the making, but they conceptualise AI as the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing, speech recognition, and machine vision.

The emergence of Gen-AI technologies expanded significantly on existing AI technologies, and opened numerous opportunities as well as future challenges. As part of ML Gen-AI uses algorithms that depend on large datasets. Current Gen-AI models and tools can produce original material, such as text, images, music, and even complex designs, based on their training data. Floridi (2024) posits that these innovations drastically change the supply and demand of information.

The AI Life Cycle

To understand AI ethics, the life cycle of AI must be considered. It is not a process that should be regarded as a separate phase, as Ng, Kapur, Blizinsky, & Hernandez-Boussard (2022:2248) claim that it must be examined with due cognisance of their impact in concert and the interactions between the phases.

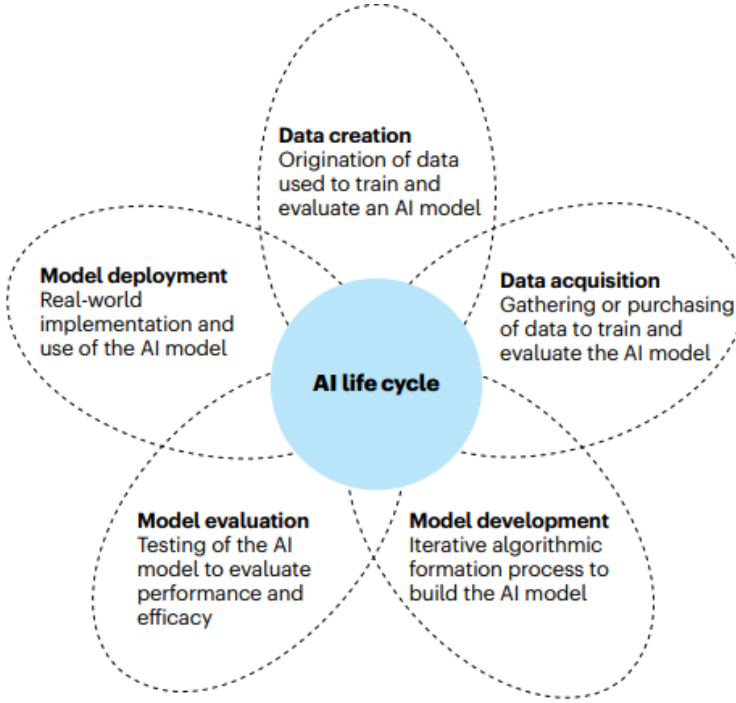

Figure 4.2: The AI life cycle. (Source: Ng et al. 2022:2247)

The AI Boom

The remarkable expansion and advancement observed in the field of AI is commonly referred to as the 'AI boom.' The swift progress in AI technology heightened corporate pursuits, and the expanded utilisation of computer capabilities in industry have contributed to enhanced customer experiences. Consequently, there has been a surge in funding towards AI technologies and the incorporation

of AI solutions into various goods and services. It is important to acknowledge that the AI industry's rapid growth is accompanied by various obstacles and factors to consider, such as ethical dilemmas, the requirement for responsible AI advancement, and the resolution of potential biases within AI systems. Furthermore, like every technological advancement, there could be phases of heightened enthusiasm followed by adaptations and development in the industry. The rate of AI progress and its influence on society will probably remain a topic of debate and exploration.

AI Ethics

Essentially, AI ethics is a system of moral principles and techniques intended to inform the development and responsible use of AI technology (Bartneck *et al.* 2020:2247). Han (2022:2 of 11) reminds us that there is no clear and agreed definition of AI, as found in other fields of ethics such as computer ethics and IE. In literature there are overlaps in terminology and often AI ethics and robotic ethics are used interchangeably (cf. Figure 3). As AI has become integral to product platforms that incorporate ethical considerations and values in the development and deployment of AI and services, organisations are starting to develop AI codes of ethics. Han (2022:3 of 11) refers to Moor (1985:266) who defines computer ethics as an analysis of the nature and social impact of computer technology and the corresponding formulation and justification of policies for the ethical use of such technology (Han 2022:3 of 11). Hauptman who was the first to use the term 'IE' in 1988, already addressed principles of censorship, privacy, access to information, balance in collection development, copyright, fair use, and codes of ethics (Froehlich 2000:265).

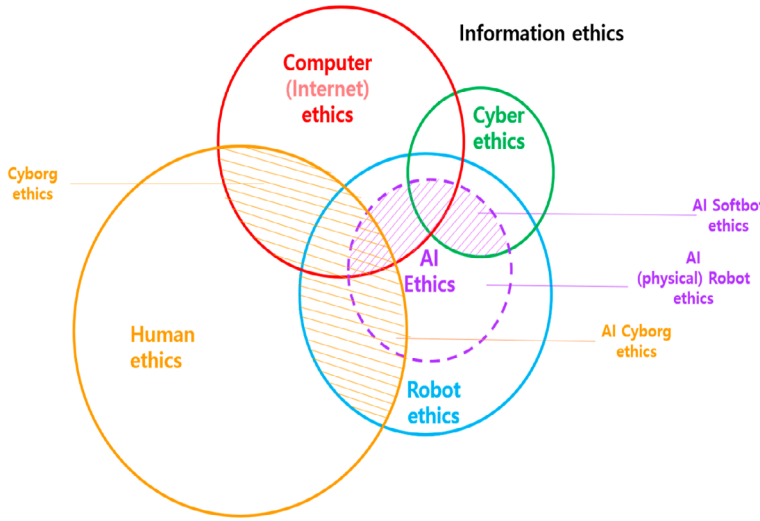

Figure 4.3: The multidisciplinary field of AI ethics. (Source: Han 2022:5 of 11)

In this model, cyborg – short for cyborg organism – and cyborg ethics refer to a combination of biological and artificial components. These enhanced capabilities are often present in biomedical technologies such as pacemakers, implants, prosthetics, or other technological enhancements that merge with the biological components of the organism. Han's model provides a contemporary view of the span and complexity of new IE.

An AI code of ethics, also called an AI value platform, is a policy statement that formally defines the role of AI as it applies to the development and wellbeing of humans. This term should not be confused with AI value-driven platforms, looking at capabilities and benefits of systems. The purpose of an AI code of ethics is to provide stakeholders with guidance when faced with an ethical decision regarding the use of AI (Ibiricu & Van der Made 2020:396). An AI value platform is generally understood as technology purposed to add value *via* AI. This could involve a combination of software, hardware, and services designed to provide solutions that address specific business challenges or opportunities.

It is important to note that the interpretation of the AI value platform may vary depending on the context and the specific goals of the entity using or providing such a platform. Organisations developing AI solutions are often looking for platforms that not only facilitate the technical aspects of AI but also deliver measurable values in terms of efficiency, innovation, and competitive advantage. For the latest and most accurate information, it is recommended to check specific industry announcements, publications, or product documentation.

It is quite incidental how Asimov's science fiction short stories around 1950 referred to an ethical code where the primary law prohibits robots from engaging in any actions that cause harm to humans or from neglecting to act when harm could be prevented (Asimov 1984:9). The second law mandates that robots must comply with human commands, except when those commands conflict with the first law. The third law mandates that robots prioritise self-preservation, if it aligns with the principles outlined in the previous two laws.

AI Ethics Frameworks and Principles

AI ethics must consider the ethical implications and societal impacts of AI technologies. Building on existing IE frameworks, new theoretical frameworks have been proposed to guide discussions and decision-making in this domain. These include the IEEE (Institute of Electrical and Electronics Engineers) global initiative on ethics of autonomous and intelligent systems and the COMEST (Commission mondiale d'éthique des connaissances scientifiques et des technologies) rights-based model: The universal declaration on ethical considerations regarding artificial intelligence and autonomous systems.

The Asilomar Principles

In 2017, a high-level conference called, *Future Life Conference* was held in California (Morandín-Ahuerma 2023:6), resulting in the Asilomar AI principles, consisting of 23 guidelines for the research and development of AI outlined developmental issues, ethics, and

guidelines for the development of AI, with the goal of guiding the development of beneficial AI.

The 23 principles were developed by a group of AI researchers, robotics, technology experts, and legal scholars from different universities and organisations. These experts organised the AI principles at the Asilomar conference while discussing the future of AI and responsible AI regulation. While these principles are not legally binding, they serve as a reference point for researchers, policymakers, and industry stakeholders.

At the forefront of the discussions to arrive at the principles was the agreement that ethics must guide all AI research to ensure privacy and fairness. Furthermore, AI research should not result in undirected knowledge. There was an agreement that ethical AI rests on cross-discipline cooperation. In creating a transdisciplinary AI research culture, trust and transparency will be uncalculated.

Even though the conference was attended by high-profile entrepreneurs like Elon Musk, the decision was that data and information security must have preference to competitiveness (Morandín-Ahuerma 2023:6). Over and above being transparent about failures, judicial transparency is important where AI is used in decision-making and any legal matters such as sentencing and accountability of the use of AI in legal systems. The risk of autonomous AI systems was realised, and the conference agreed that there should always be an alignment with human values and the dichotomy of freedom versus privacy in terms of the use of private data, and that regulations should be such that individuals are not unduly restricted to access. The conference and subsequent decisions stressed the importance of human control. The delegation of these types of autonomous tasks must reside with humans. The conference addressed the future capabilities of AI that should be regulated and that the future of the planet should be the guiding principle.

A critically important discussion at the conference was about AI and warfare. It was emphasised that human life must be valued and that AI armoury should be designed with great care. This must be done with international cooperation and

transparency. The conference gave due consideration to the dangers of AI, and the importance to be vigilant and plan for possible risks. In conclusion, the main agreement was that AI must be developed and used for the greater good.

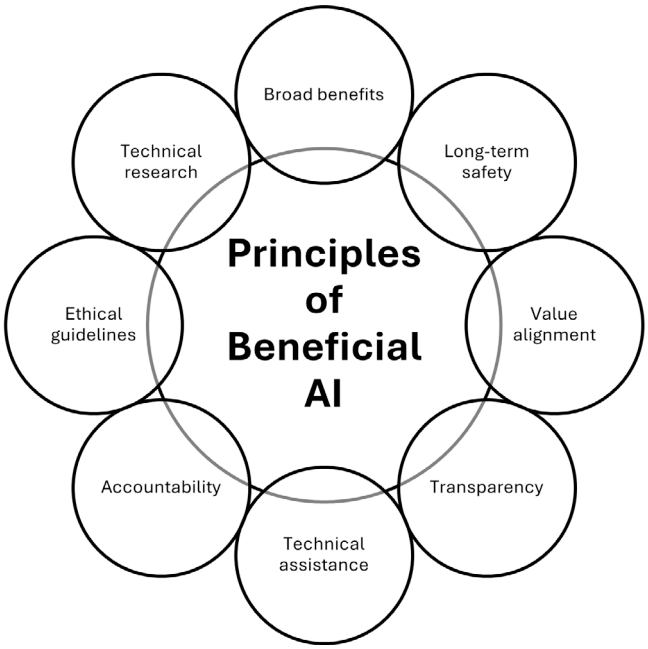

Figure 4.4: Principles of beneficial AI. (Source: Adapted from Morandín-Ahuerma 2023:20-24)

It is worth noting that these principles represent a consensus among the participants of the Asilomar conference at the time, as the discussions on AI ethics and guidelines have continued to evolve since then. Various organisations and initiatives around the world continue to contribute to the development of ethical frameworks for AI. In addition to the principles mentioned in Figure 4, trust and fairness in AI, AI and social biases, and social justice and cultural sensitivity are elements of ethical importance.

Trust and Fairness in AI

When discussing AI ethics, the topic of algorithm ethics invariably emerges. Algorithms are knowingly or unknowingly part of many aspects of life (Tsamados, Aggarwal, Cowls, Morley, Roberts, Taddeo, & Floridi 2021). Think of the recommender systems of streaming services or online purchases, all based on algorithms. Benk, Tolmeijer, Von Wangenheim, & Ferrario (2022:2 of 12) underpin the importance of creating trust in AI-based systems towards improved ethical and effective application and use. They continue to lament that the paucity in literature, a shared understanding, and measurement of what trust within the context of AI means, have negatively impacted progress in the field (Benk *et al*. 2022:2 of 12). Decision-making algorithms are increasingly being used in industry, of which legal decisions and healthcare are at the forefront (Vaassen 2021:1 of 20). A growing number of authors (cf. e.g., Burrell, 2016:3 of 12; Benk *et al*. 2022:2 of 12) refer to the black box concept to explain the lack of transparency and interpretability in certain advanced machine learning models, especially with deep neural networks. In this context, the term 'opacity' indicates that there are instances when the outcomes of an algorithm and how it was derived are not clear (Vaassen 2021:2 of 20). The opacity of black-box algorithms, particularly in the context of ML models, can give rise to various problems and challenges (cf. Figure 4.5).

It is furthermore possible that bias can transpire which will influence decisions (Burrell 2016:2 of 12; Vaassen 2021:3 of 20). Establishing accountability and responsibility for algorithmic outcomes becomes challenging, which may have legal and ethical implications.

Black-box algorithms may inadvertently learn and perpetuate biases present in the training data. If unchecked, discriminatory outcomes may occur, reinforcing or exacerbating existing societal biases. Users may be reluctant to rely on or accept algorithmic decisions if they cannot understand or validate the decision-making process.

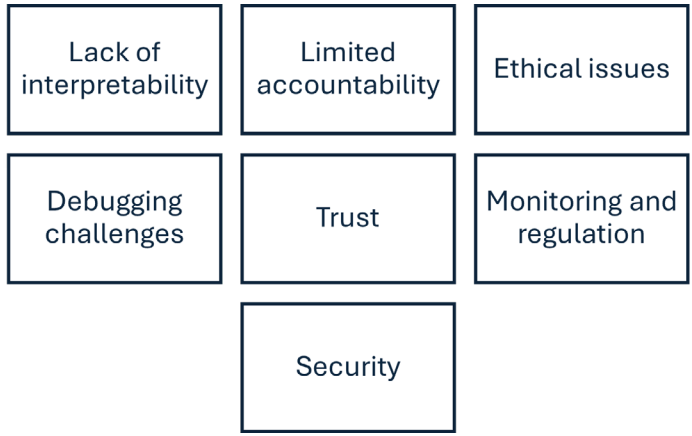

Figure 4.5: Opacity in black box algorithms. (Source: Based on Burrell 2016)

It stands to reason that areas with critical societal impact such as financial, health, and legal aspects require firm regulatory frameworks and call for more clarity on explainability, transparency, and trustworthiness of AI applications.

Algorithm Ethics

With obvious tangents and overlapping with AI ethics, algorithm ethics refers to the ethical considerations associated with the design, development, deployment, and impact of algorithms. Algorithms are sets of instructions or rules followed by a computer programme to perform a specific task or solve a particular problem. Algorithms can inadvertently perpetuate biases present in training data, leading to unfair or discriminatory outcomes (Vaassen 2021:5 of 20). Ethical algorithm design aims to minimise bias, ensure fairness, and prevent discrimination, especially in areas like hiring, lending, and law enforcement.

Whereas algorithms process sensitive personal information, it raises concerns about privacy infringements. Ethical algorithm development involves implementing privacy-preserving measures to protect individuals' data and ensuring compliance with data protection regulations. Transparent and explainable algorithms enhance accountability, trust, and understanding,

allowing users to comprehend and channel decisions. Identifying the responsibility for algorithmic outcomes can be challenging, especially in complex systems. Ethical algorithms include mechanisms for accountability, making it clear who is responsible for the decisions made by the algorithm and providing avenues for redress in case of errors. Algorithms may be vulnerable to malicious attacks or manipulation. Ethical algorithms involve transparent communication with users, ensuring they understand how algorithms impact them and giving them control over their data and preferences. Ethical algorithms prioritise inclusivity and accessibility, aiming to minimise biases and ensure equitable outcomes for diverse populations. Ethical algorithms include mechanisms for continuous monitoring, evaluation, and improvement to ensure their ongoing effectiveness and fairness.

AI Biases

AI is prone to a range of biases, of which social biases is of particular concern. Friedman and Nissenbaum (1996:332) define three types of bias: Preexisting bias based on social practices and attitudes likely embedded in data used for 'training;' technical bias based on design constraints in hardware and software likely in AI to be associated with limits of available data and a lack of qualitative nuance; and emergent bias arising from changing the used context such that accurate meaning is lost or distorted when viewed through a different prism. The potential of bias to permeate at least some categories of AI applications is a significant concern (Niederman & Baker 2022). Other AI biases include algorithm biases and data biases.

AI, Social Justice, and Cultural Sensitivity

AI ethics in the medical industry primarily focuses on the objective of guaranteeing impartiality and preventing biases in AI-powered decision-making. Given the varied patient demographics and the importance of fair treatment, this is of utmost importance. AI systems in healthcare necessitate careful design and ongoing monitoring to mitigate the impact of inherent biases in training data on diagnostic or treatment suggestions. This is also connected to the wider problem of digital exclusion, where

marginalised or underrepresented groups may not have equitable access to AI-enhanced medical services, therefore worsening existing healthcare inequalities.

Within the realm of law, the utilisation of AI implies a distinct array of ethical dilemmas. AI systems have the power to aid in legal research, analyse documents, and even make predictions about case outcomes. However, they also give rise to significant concerns around transparency, accountability, and fairness in judicial proceedings. AI utilisation in sentencing or parole determinations need a meticulous regulation to guarantee that algorithmic suggestions do not perpetuate past prejudices or violate legal rights.

Digital Exclusion

Digital exclusion involves not only the lack of access to technology, but also the inability to comprehend and engage with digital systems. With the increasing integration of AI into daily life, there is a risk that certain groups may face a disproportionate disadvantage due to the digital divide, as they may lack the necessary resources or expertise to fully utilise the breakthroughs in AI.

Ethical intelligence in AI pertains to the creation of AI systems that not only exhibit efficient task performance but also adhere to ethical and moral principles. This task is especially difficult due to the subjective character of ethics and the intricate process of programming machines to comprehend human values.

AI Risks

In the dynamic world of technology, ethical risks of AI do not only feature in its use, but also in its design. Addressing challenges of biases in data, algorithms, as well as training set limitations leaves room for deeper discourse and ongoing efforts to improve data quality, enhance algorithmic fairness, ensure robustness, and consider ethical implications throughout the AI development lifecycle (cf. Figure 4.2).

AI could be prone to inaccurate information and misrepresentation. Some of these challenges include incidences of deep fake, identity theft, and privacy issues in using facial recognition, to name but a few.

Deep Fake

The utilisation of AI to generate authentic yet fabricated audio or video content, known as deepfake technology, poses a significant ethical quandary. Although it can be used for harmless purposes such as amusement, it also has the capacity to be misused for purposes such as disseminating false information, impersonating others, and infringing upon privacy and permission.

AI and Facial Recognition

Facial recognition technology, a controversial form of AI, has advantages in terms of security and customisation, but also gives rise to significant worries regarding privacy, monitoring, and the possibility of misuse. The ethical use of these technologies is further complicated by misidentification and prejudices, notably towards specific ethnic groups.

Identity Theft

Ahmed (2020:2) alludes that identity theft is on the rise but warns that this crime is still ill-defined. In essence it entails the wrongful acquirement and utilisation of another person's personal information, such as their name, social security number, credit card information, or other identifying details, thus without their permission. The goal of identity theft is typically to commit fraud, financial theft, or other criminal activities using the victim's identity. Some of the methods include phishing, data breaches, dumpster diving, manipulative social engineering, and skimming of credit cards. Most identity crimes take place where people illegally obtain information by purchasing it from others, skilled in accessing and using information online (Ahmed 2020:193). These crimes cause tremendous financial, emotional, and at times, reputational damage.

AI Threats

A dynamic invention like AI invariably poses some threats to society. Much of these lies in the misuse of AI like cyberwarfare, and there is also the overarching fear of loss of human control when autonomous systems move closer to super-intelligence.

The Threat of Singularity

The term 'singularity' was first coined by John von Neumann in 1948 (Ulam 1958). Singularity, as described by scholars such as Shanahan (2015), posits a future moment when AI systems attain a state of independent functioning as a result of the advent of machine awareness and superhuman intellect. This concept imagines a situation in which AI exceeds human cognitive capabilities, resulting in a revolutionary and perhaps unmanageable stage in technology development.

Singularity gives rise to significant ethical, intellectual, and existential inquiries. The main apprehension revolves around the possible relinquishment of human authority over technology. The inherent uncertainty of AI systems presents substantial hazards, such as the potential for destructive actions that may harm humans or the environment. The issue of whether the singularity is presently in progress is a subject of contention among technologists, ethicists, and futurists. The swift progress in AI technology implies that the possibility of singularity, although hypothetical, cannot be completely disregarded.

The concept of singularity has a complex and diverse impact on AI ethics and poses a challenge to our existing comprehension of ethical frameworks in a society where AI possesses autonomous decision-making abilities, as the task of assigning accountability for the consequences of those actions, whether they are advantageous or detrimental, becomes a multifaceted matter.

Ethics of Gen-AI

It should be noted that the technology of Gen-AI is not new. Gen-AI was introduced in the 1960s in chatbots. However, it was not until 2014, with the introduction of GANs (generative adversarial

networks) – a type of ML algorithm – that Gen-AI could create convincingly authentic images, videos, and audio of real people. Gen-AI is a type of AI technology that can produce various types of content, including text, imagery, audio, and synthetic data. Gen-AI refers to a class of AI systems that can generate new content, such as images, text, or even music, that is not explicitly programmed. These systems are designed to understand and learn patterns from existing data and then use that understanding to create entirely new data that resemble the training set. Bandi, Adapa, and Kuchi (2023:1 of 60) explain that Gen-AI is geared to develop algorithms able to create synthetic data. There are several types of Gen-AI models, and one of the most notable is the class of generative models based on neural networks.

Gen-AI's ability to generate data resembling real-life conditions, allows for content creation (Bandi *et al.* 2023:1 of 60). It is a further development of Web 3.0. Gen-AI has found applications in various domains, including text generation in NLP (natural language processing) models, such as GPT 3 (generative pre-trained transformer 3) and GPT 4 and data augmentation of data training sets, and can generate coherent and contextually relevant text. These models can be used for tasks like writing articles and generating creative content. Then there is the image generation, where generative models can create realistic images that resemble photographs of human faces, animals, or scenes.

While Gen-AI has shown remarkable capabilities, there are also ethical concerns, especially regarding the potential misuse of the technology, such as the creation of deepfakes for malicious purposes. As a result, ongoing research and development in the field aim to address both the positive and negative implications of Gen-AI.

Large Language Models and Gen-AI Ethics

There are overlaps and tangents in ethical concerns in the use and applications of Gen-AI, but the nature of LLMs adds an element of further concern – that of bias and privacy. LLMs refer to advanced AI models that are specifically designed for NLP tasks.

The primary architecture used for LLMs is the transformer architecture, which are neural networks. The transformer architecture allows for parallelisation and efficient processing of sequences, making it well-suited for handling natural language. Transformer models apply an evolving set of mathematical techniques, called attention or self-attention, to detect subtle ways. Even distant data elements in a series influence and depend on each other. LLMs have analytical abilities, which can be used for chatbots and virtual agents. The transformative nature of large language models lies in their ability to fundamentally change how we interact with and leverage natural language in various fields, impacting communication, creativity, productivity, and research. These LLMs are often pre-trained on massive datasets and fine-tuned for specific tasks. The use of large-scale language models harbours several risks and ethical concerns (Weidinger, Mellor, Rauh, Griffin, Uesato, Huang, Cheng, Glaese, Balle, Kasirzadeh, Kenton, Brown, Hawkins, Stepleton, Biles, Birhane, Haas, Rimell, Hendricks, Isaac, Legassick, Irving, & Gabriel 2021:7 of 64). Ethical concerns in the use of LLMs include misinformation and disinformation, perpetuating bias, a lack of explainability, the risk of manipulation, as well as privacy and security concerns.

LLMs and Training Data

A LLM model is a statistical model that predicts the probability of a sequence of words. Based on DL (deep learning) – a branch of ML – LLMs import vast amounts of data, openly available from books, web pages, or similar contents (Paass & Giesselbach 2023:21). Then patterns are identified to get connections between words. A LLM with a large feed of data can generate content better, faster, and more accurately, which then generates new content based on the prompts provided by the user. Training data in LLMs involve exposing the model to vast amounts of diverse text from a wide range of sources.

It is important to note that the success of LLMs like GPT 3 is attributed to their ability to generalise from diverse training data. The models can then generate coherent and contextually relevant text across a wide range of topics and tasks. The ethical considerations related to the use of such models, including biases

in training data and potential societal impacts, are also important aspects to consider in the development and deployment of these models.

Table 4.2: Summary of Gen-AI ethical challenges (Adapted from Weidinger et al. 2021)

LLM Ethical Concerns	**LMM May Predict as Follows**
Discrimination	Social stereotyping based on gender, religion, orientation, ability, and age
Toxicity and exclusion Unjust, prejudiced, oppressive use of natural language in models	
Harm caused by misinformation	False, misleading, poor-quality information
Intentional malicious use	Use LMM for illegal surveillance, fraud, scams, deepfake, identity theft, and censorship
Information hazards	LMM predictions include privacy and safety risks – leaking private information
Harm from human-computer interaction	Dependency, too trusting, psychological profile and vulnerability, ethnic profiling
Environmental harm	Economic disparity, energy, and water demands
Access	Disparate access, exclusion, literacy, and skills constraints

LLMs like ChatGPT generate responses based on patterns learned from large datasets (training data), and at times they may generate content that appears to be contextually appropriate, although it is not factually correct. This is referred to as AI hallucination and is a growing ethics concern. While LLMs can be powerful tools for generating human-like text, they do not have a genuine understanding or awareness yet.

AI programmes, more specifically LLMs, have the potential to provide deceptive, erroneous, or wholly fictional outputs. These present notable ethical dilemmas, particularly when such technologies are employed for crucial decision-making or distributing information. To mitigate hallucinations and improve the reliability of LLMs, R&D (research and development) efforts

focus on refining training methodologies, incorporating diverse datasets, addressing biases, and enhancing the model's ability to handle nuanced and contextually sensitive information.

Intergovernmental and other Initiatives in Support of New Dimensions in Dynamic IE Ethics Innovation

Floridi (2014:218) urges that any information society must be equipped with sustainable IE, and that this be made well known. It is here where intergovernmental organisations have a critical role to play, moreover in the time where disruptive technologies could potentially add to chaotic and unethical consequences. An example of the speed at which innovation takes place is the rapid development of the Gen-AI by OpenAI in bringing out its chatbot.

UNESCO has made great strides in addressing frontier challenges in IE, the ethics of AI, and the ethics of neurotechnology. In addition, they are addressing issues on climate engineering and the IoT (internet of things), which can no longer be separated from IE and ethics of AI. Neuroethics is an emerging field. Combined with AI these techniques can enable developers – public or private – to abuse cognitive biases and trigger reactions and emotions without consent (Farisco, Salles, & Evers 2018:718). Whereas neural networks in AI are concerned with the development and application of algorithms for pattern recognition, classification, regression, and other task, neuroscience is a multidisciplinary fields of study that examines the ethical, legal, and social implications of neuroscience. Essentially, it encompasses the scientific study of the nervous system that regulates all human cognition, behaviour, and functions. The field encompasses a broad range of topics, from understanding how individual neurons function to exploring complex behaviours and cognitive processes. However, there is a link to AI because the more people discover how the nervous system works in relation to our world, the more there will be neuro-AI tangents (Berger & Rossi 2022:2054) together with the encompassing ethical challenges. An example of neuroscience and AI is wearable devices that integrate technologies from both neuroscience and AI to monitor, analyse, or interact with the brain

and its functions. These wearables often serve various purposes, including research, healthcare, and personal wellbeing.

Based on their ability to collaborate across barriers such as borders, language, and cultural differences, organisations with the value and importance such as UNESCO, the OECD, and others are critical in researching and guiding society and the world in the ethical deployment of AI technologies. UNESCO emphasises the importance of promoting inclusivity in AI deployment for enhanced respect of cultural diversity. For this they provide ethical guidelines to address issues such as transparency, accountability, fairness, and the impact of AI on human rights. They advocate better education and capacity building specifically in the fields of IE and ethics of AI. However, their foremost value lies in the opportunities created to foster international cooperation and dialogue on AI ethics. This involves collaboration with other international organisations, governments, academia, industry, and civil society to develop shared principles and norms.

In the 1970s and 1980s, the OECD encouraged member countries to develop guidelines and legislation to manage data responsibly (Wright, De Hert, & Gutwirth 2011:119). In the seven-year stint between 1973 and 1980, one-third of the OECD's 30 member countries enacted legislation intended to protect individuals against the abuse of data related to them and to give individuals the right of access to data with a view to checking their accuracy and appropriateness. The OECD is acutely aware of the tension that exists between the need of free flow of information and protecting of personal data (Wright *et al.* 2011:120). The 2013 privacy framework by the OECD and later in 2023 the OECD's *Good practice principles for data ethics in the public sector* (OECD 2023), offer guardrails to the ethical use of information.

An Intergovernmental Organisation's Role in Fair Information Practice

As early as 1980, the OECD's privacy guidelines outlined principles for the protection of privacy and personal data (cf. OECD 2002). These principles are not legally binding, but they have influenced the development of privacy laws and frameworks around the

world. The key principles include putting limitations on the illegal collection of personal information, ensuring data quality.

These principles are widely considered foundational for the development of privacy laws and regulations. Various countries and regions have adapted and implemented these principles in their own privacy frameworks. It is essential to check the latest developments and specific regulations in one's jurisdiction, as privacy laws can vary significantly.

The OECD Fair Information Practice

The OECD FIP principles are fundamental guidelines that underpin data protection and privacy policies worldwide. The principles were formulated by the OECD in the 1980s with the aim of addressing apprehensions regarding the collection, processing, and utilisation of personal information. The FIP principles include important elements such as restricting data collecting to essential and legal purposes, guaranteeing data accuracy, defining data security rules, and promoting a dedication to openness and transparency in data practices. In addition, they emphasise the significance of individual engagement, affording individuals the privilege to obtain and rectify their personal information. These principles have had a significant impact on many national and international laws and recommendations regarding data protection. They have provided the basis for modern privacy rules such as the GDPR (general data protection regulation) in the EU (European Union). The relevance of the OECD FIP principles persists in the present day, as they offer a structure for managing the progress of data processing technology while safeguarding individual privacy rights.

Pointers for Responsible AI

Table 4.3 below is based on an overview of literature, where specific deductions on essential elements that will ensure responsible AI use and development must be considered.

Table 4.3: Elements of responsible AI development and use

Element	Use
Ethics frameworks and standards	Establish and comply with ethical principles that regulate the development and implementation of AI, guaranteeing that AI systems are created and utilised in a way that upholds human rights, dignity, and ethical norms.
Transparency and explainability	Guarantee that AI systems are clear and comprehensible in their operations and decision-making processes. Create AI models that are transparent and comprehensible to people, fostering increased trust and accountability.
Fairness and non-discrimination	Promote equity and impartiality by proactively addressing and minimising biases in AI systems. Conduct frequent audits and testing of AI algorithms to verify that they do not perpetuate or worsen discrimination based on race, gender, age, or other attributes.
Privacy and data protection	Ensure the implementation of strong data protection procedures to preserve personal and sensitive information. Ensure the preservation of user privacy and adhere to applicable data protection legislation.
Accountability and responsibility	Define explicit channels of responsibility for the decisions and acts of AI systems. Establish safeguards to provide recourse if AI systems inflict harm or function in unanticipated manners.
Safety and security	Emphasise the primacy of ensuring the physical and digital security of AI systems. Safeguard against the malevolent exploitation of AI technology and guarantee the robustness of systems against hacking and other cyber threats.
Sustainability and environmental impact	Evaluate the ecological ramifications of AI systems. Strive for the development and implementation of AI that is energy-efficient to reduce carbon emissions and support sustainability.
Inclusivity and accessibility	Ensure inclusivity and accessibility by designing AI systems that can easily be accessed and used by all individuals. Consider the varied requirements and capabilities of all possible users and strive to ensure that AI technology is advantageous and accessible to a broad range of individuals.

Element	Use
Collaboration and engagement	Interact with stakeholders, such as policymakers, industrial partners, academic researchers, and the public, to comprehend various viewpoints and tackle societal issues associated with AI.
Monitoring, improvement, regulation, and collaboration	Consistently assess and enhance AI systems to guarantee their adherence to ethical norms, legal obligations, and social values. Keep oneself updated on the latest developments in AI and adapt one's practices accordingly.
Education, literacies, and awareness	All educational institutions have a responsibility to develop composite literacies and awareness around the nature of AI and the responsible use thereof.

Following these guidelines helps to guarantee the responsible and ethical development and use of AI technology, with the aim of benefiting society while minimising any dangers and negative consequences.

Conclusion

To summarise, the field of AI ethics and its implementation in other areas like medicine and law, pose an intricate and dynamic challenge. The discussion about AI and the possibility of singularity, along with the implementation of fair information practices and responsible AI standards, highlight the need for a comprehensive approach to the development and regulation of AI.

We are currently at a critical point in time where the swift progress in AI technology, as demonstrated by systems such as GPT 4 and the expected development of models like GPT 5, necessitate a proactive and deliberate approach to ethical considerations. The OECD FIP standards serve as a fundamental framework for safeguarding data and ensuring privacy, which are of utmost importance in the era of AI. These principles, in conjunction with the guidelines for responsible AI, provide a clear path for the ethical, transparent, fair, and inclusive development of AI.

As we progress deeper into this era of technology, it is crucial to maintain a balance between innovation and accountability. AI development must be guided by ethical frameworks and norms to

ensure that these technologies are utilised for the betterment of society while upholding human rights and dignity. Transparency, explainability, and accountability are not mere aspirations, but rather crucial prerequisites for fostering confidence and the approval of AI systems. To mitigate hazards connected with AI, it is necessary to address concerns of justice and non-discrimination, prioritise privacy and data protection, and ensure safety and security.

Furthermore, the possible consequences of AI attaining or exceeding human cognitive capabilities – a central concept in the singularity discourse – emphasise the pressing need for these ethical deliberations. Continuous monitoring, stakeholder interaction, and adaptation is crucial in ensuring ethical practices in AI. The advancement of AI, with its potential benefits and difficulties, necessitate a cooperative endeavour involving technologists, ethicists, legislators, and the broader society. By engaging in such partnerships, we can guarantee that the development made in AI is in accordance with ethical values and has a good impact on the advancement of humanity.

Therefore, as we conclude this chapter, it is evident that the progression of AI ethics is continuous. The concepts and methods covered here are dynamic, continuously adapting to the technology they endeavour to regulate. The task at hand is not alone to responsibly advance AI, but also to consistently modify our ethical frameworks to align with the always evolving AI technology and its influence on society.

References

Alharbi, M. & Mukhari, A. 2023. Information science and interdisciplinary: Literature review. *International Journal of Advances in Science Engineering and Technology* 9(2):38-42.

Ahmed, SR. 2020. *Preventing identity crime: Identity theft and identity fraud: An identity crime model and legislative analysis with recommendations for preventing identity crime*. Leiden: Brill. https://doi.org/10.1163/9789004395978

Asimov, I. (Ed.). 1984. *Isaac Asimov presents the great SF stories 12 (1950)*. New York: DAW Books.

Bandi, A., Adapa, PVSR., & Kuchi, YEVPK. 2023. The power of generative AI: A review of requirements, models, input-output formats, evaluation metrics, and challenges. *Future Internet* 15, 260. 60 pages. https://doi.org/10.3390/fi15080260Bandi

Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. 2020. *An introduction to ethics in robotics and AI*. Cham: Springer. https://doi.org/10.1007/978-3-030-51110-4

Bates, MJ. 2005. Information and knowledge: An evolutionary framework for information science. *Information Research* 10(4). 239. 30 pages.

Benk, M., Tolmeijer, S., Von Wangenheim, F., & Ferrario, A. 2022. The value of measuring trust in AI – a socio-technical system perspective. *arXiv:2204.13480v1*. 12 pages. Available at: https://arxiv.org/pdf/2204.13480.pdf. (Accessed on 23 January 2024).

Berger, SE. & Rossi, F. 2022. Addressing neuroethics issues in practice: Lessons learnt by tech companies in AI ethics. *Neuron* 110(13):2052-2056. https://doi.org/10.1016/j.neuron.2022.05.006

Bester, C. & Fischer, R. 2020. The essential relationship between information ethics and artificial intelligence. *Artificial Intelligence, Ethics and Society* 29:1-11. https://doi.org/10.29173/irie428

Birrer, FAJ. 1999. Understanding values and biases in IT. *ACM SIGCAS Computers and Society* 29(1):16-21. https://doi.org/10.1145/382042.382047

Borko, H. 1968. Information science: What is it? *Journal of the Association for Information Science and Technology* 19(1):3-5. https://doi.org/10.1002/asi.5090190103

Burgess, JTF. & Knox, EJM. (Eds.). 2019. *Foundations of information ethics*. Chicago: American Library Association.

Burrell, J. 2016. How the machine 'thinks:' Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1). 12 pages. https://doi.org/10.1177/2053951715622512

Bynum, TW. 2010. Philosophy in the information age. *Metaphilosophy* 41(3):420-442. https://doi.org/10.1111/j.1467-9973.2010.01651.x

Capurro, R., Eldred, M., & Nagel, D. 2013. *Digital whoness: Identity, privacy and freedom in the cyberworld*. Frankfurt: Ontos Verlag. https://doi.org/10.1515/9783110320428

Ess, C. 2008. Luciano Floridi's philosophy of information and information ethics: Critical reflections and the state of the art. *Ethics and Information Technology* 10(2-3):89-96. https://doi.org/10.1007/s10676-008-9172-8

Ess, C. 2014. *Digital media ethics*. Revised and updated 2nd ed. Cambridge: Polity.

Floridi, L. 1999. Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology* 1(1):33-52. https://doi.org/10.1023/A:1010018611096

Floridi. L. 2010. *The Cambridge handbook of information and computer ethics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511845239

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

Floridi, L. 2014. *The fourth revolution: How the infosphere is reshaping human reality*. Oxford: Oxford University Press.

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

Floridi, L. 2016. Mature information societies – a matter of expectations. *Philosophy & Technology* 29(1):1-4. 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. 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. *Philosophy & Technology* 33:1-3. https://doi.org/10.1007/s13347-020-00396-6

Floridi, L. 2024. On the future of content in the age of artificial intelligence: Some implications and directions. *Philosophy & Technology* 37. 112. 11 pages. https://doi.org/10.1007/s13347-024-00806-z

Farisco, M., Salles, A., & Evers, K. 2018. Neuroethics: A conceptual approach. *Cambridge Quarterly of Healthcare Ethics* 27(4):717-727. https://doi.org/10.1017/S0963180118000208

Friedman, B. & Nissenbaum, H. 1996. Bias in computer systems. *ACM Transactions on Information Systems (Tois)* 14(3):330-347. https://doi.org/10.1145/230538.230561

Froehlich, TJ. 2000. Intellectual freedom, ethical deliberation and codes of ethics. *IFLA Journal* 26(4):264-272. https://doi.org/10.1177/034003520002600405

Han, J. 2022. An information ethics framework based on ICT platforms. *Information* 13. 440. 11 pages. https://doi.org/10.3390/info13090440

Ibiricu, B. & Van der Made, ML. 2020. Ethics by design: A code of ethics for the digital age. *Records Management Journal* 30(3):395-414. https://doi.org/10.1108/RMJ-08-2019-0044

Mason, RO. 1986. Four ethical issues of the information age. MIS Quarterly 10(1):5-12. https://doi.org/10.2307/248873

Moor, JH. 1985. What is computer ethics? *Metaphilosophy* 16(4):266-275. https://doi.org/10.1111/j.1467-9973.1985.tb00173.x

Moor, JH. 2006. The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems* 21:18-21. https://doi.org/10.1109/MIS.2006.80

Moor, JH. 2020. What is computer ethics? In Miller, KW. & Taddeo, M. (Eds.): *The ethics of information technologies*, 15-24. London: Routledge. https://doi.org/10.4324/9781003075011-1

Morandín-Ahuerma, F. 2023. Twenty-three Asilomar principles for artificial intelligence and the future of life. *OSF Preprints* 5-27. https://doi.org/10.31219/osf.io/dgnq8.

Ng, MY., Kapur, S., Blizinsky, KD., & Hernandez-Boussard, T. 2022. The AI life cycle: A holistic approach to creating ethical AI for health decisions. *Nature Medicine* 28(11):2247-2249. https://doi.org/10.1038/s41591-022-01993-y

Niederman, F., & Baker, EW. 2022. Ethics and AI issues: Old container with new wine? *Information Systems Frontiers: A Journal of Research and Innovation* 25(1):9-28. https://doi.org/10.1007/s10796-022-10305-1

OECD (Organisation for Economic Cooperation and Development). 2002. OECD guidelines on the protection of privacy and transborder flows of personal data. Available at: https://www.oecd.org/en/publications/oecd-guidelines-on-the-protection-of-privacy-and-transborder-flows-of-personal-data_9789264196391-en.html. (Accessed on 23 March 2024).

OECD (Organisation for Economic Cooperation and Development). 2013. The OECD privacy framework. Available at: https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf. (Accessed on 30 November 2023).

OECD (Organisation for Economic Cooperation and Development). 2023. Good practice principles for data ethics in the public sector. 10 August 2023. Available at: https://digital-skills-jobs.europa.eu/en/inspiration/resources/oecds-good-practice-principles-data-ethics-public-sector. (Accessed on 23 March 2024).

Paass, G. & Giesselbach, S. 2023. *Foundation models for natural language processing: Pre-trained language models integrating media*. Cham: Springer. https://doi.org/10.1007/978-3-031-23190-2

Quinn, MJ. 2011. *Ethics for the information age*. 4th ed. New York: Pearson.

Sartori, L. & Theodorou, A. 2022. A sociotechnical perspective for the future of AI: Narratives, inequalities, and human control. *Ethics and Information Technology* 24(4). 11 pages. https://doi.org/10.1007/s10676-022-09624-3

Shanahan, M. 2015. *The technological singularity*. Cambridge: MIT Press. https://doi.org/10.7551/mitpress/10058.001.0001

Shannon, CE. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3):379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Solove, DJ. 2006. A taxonomy of privacy. *University of Pennsylvania Law Review* 154(3):477-560. https://doi.org/10.2307/40041279

Solove, D. 2015. The growing problems with the sectoral approach to privacy law. *Privacy + Security Blog*. 13 November 2015. https://teachprivacy.com/problems-sectoral-approach-privacy-law. (Accessed on 13 November 2024).

Stahl, BC. 2008. Discourses on information ethics: The claim to universality. *Ethics and Information Technology* 10:97-108. https://doi.org/10.1007/s10676-008-9171-9

Strickland, E. 2021. The turbulent past and uncertain future of AI: Is there a way out of AI's boom-and-bust cycle? *IEEE Spectrum* 58(10):26-31. https://doi.org/10.1109/MSPEC.2021.9563956

Tavani, HT. 1999. Informational privacy, data mining, and the internet. *Ethics and Information Technology* 1(2):137-145. https://doi.org/10.1023/A:1010063528863

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. 2021. The ethics of algorithms: Key problems and solutions. *AI & Society* 37:215-230. https://doi.org/10.1007/978-3-030-81907-1_8

Turing, AM. 1950. Computing machinery and intelligence. *Mind* 59:433-460. https://doi.org/10.1093/mind/LIX.236.433

Ulam, S. 1958. Tribute to John von Neumann. *Bulletin of the American Mathematical Society* 64(3):1-49. https://doi.org/10.1090/S0002-9904-1958-10189-5

Vaassen, B. 2021. AI, opacity, and personal autonomy. *Philosophy & Technology* 35. 88. 20 pages. https://doi.org/10.1007/s13347-022-00577-5

Wilson, TD. 1997. Information behaviour: An interdisciplinary perspective. *Information Processing & Management* 33(4):551-572. https://doi.org/10.1016/S0306-4573(97)00028-9

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, LA., Isaac, W., Legassick, S., Irving, G., & Gabriel, I. 2021. Ethical and social risks of harm from language models. *arXiv:2112.04359v1*. 64 pages. https://doi.org/10.48550/arXiv.2112.04359

Woodward, B., Imboden, T., & Martin, NL. 2011. An undergraduate information security program: More than a curriculum. *Journal of Information Systems Education* 24(1):63-70.

Wright, D., De Hert, P., & Gutwirth, S. 2011. Are the OECD guidelines at 30 showing their age? *Communications of the ACM* 54(2):119-127. https://doi.org/10.1145/1897816.1897848

Young, J., Smith, TJ., & Zheng, SH. 2020. Call me BIG PAPA: An extension of Mason's information ethics framework to big data. *Journal of the Midwest Association for Information Systems (JMWAIS)* 2:17-42.

Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S., & Savage, S. 2020. A survey on ethical principles of AI and implementations. *IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, ACT, Australia, 3010-3017. https://doi.org/10.1109/SSCI47803.2020.9308437