A Fresh Look at Ethical Perspectives on Artificial Intelligence Applications and their Potential Impacts at Work and on People

Eddie John Paul Fisher

Skema Business School, Lille and Paris, France

Brno University of Technology, Brno, Czech Republic

Eddie Fisher (Corresponding author)

8 Kendal, Swindon, Wiltshire, SN5 8HW, United Kingdom

Tel: 44-179-349-0423   E-mail: eddie.fisher9@btinternet.com

Abstract

Artificial Intelligence (AI) has recently attracted a heightened interest within the communities of AI theory and practice. Experts in this area have advanced their knowledge and understanding of AI systems and applications and the impact of these on existing technologies in terms of capabilities and benefits/risks. The development of these capabilities has created a different choice that needs to be decided on: let the potential effects of AI grow and become more exciting, at the expense of disturbing ethical concerns and issues, or balance these exciting developments by introducing suitable legislative approaches to address these concerns. The results from this research imply that urgent attention needs to be given to construct primary legislation (acts of Parliament, statutes) and then implement these as a matter of urgency. More focus and attention must be placed on ethical concerns and privacy issues that relate to existing and planned AI developments. It is further suggested that more collaboration and co-operation must be exercised across geographical boundaries between researchers to ensure that the interests of human beings world-wide are considered of paramount importance when developing and rolling out AI systems and applications.

Keywords: Artificial Intelligence, Human Rights, AI Regulations, Smart City, Discrimination, Privacy

1. Introduction

1.1 Introduction

Recent research purports that artificial intelligence (AI) is progressing at a fast pace. Some of its currently recognised capabilities include the transcription of speech, object recognition in videos, translations between languages, the preparation of legal documents and identification of cancer in tissues (Liao, 2020). Based on this current AI development trend, AI is not only expected to reach but to exceed human performance in more challenging and complex tasks. With the rapid advancement of AI technologies, there appears to be an urgent need to identify and raise ethical questions and concerns about current and future AI systems and applications. Many people are already using this new technology without further thought, such as the use of satellite navigation systems in their cars. As more of AI technology is developed and becomes available, people are developing increasing concerns about the latest AI developments. There appears to be a universal consensus that appropriate and effective controls need to be put in place to ensure that the raised ethical AI concerns are addressed and resolved. This research conducted a comprehensive literature review to establish what is already known about the subject matter under investigation and how the outcome could be put into practice (bringing together theory and practice much closer). The primary focus was on the following three research domains: 1 Humanistic concern such as unemployment, privacy invasion and discrimination 2 Regulatory position such as AI policy, accountability, and regulations 3 Specific purpose application such as climate change challenges, the military and healthcare.

The research objectives of this research were:

- To summarise the relevant facts about known ethical concerns.

- To determine the most salient ethical concerns.

- To encourage fresh thinking and approach to AI applications within the community of AI practice.

- To propose a list of practical recommendations how to improve and resolve the identified AI ethical concerns.

There is no definitive agreed statement of what is meant by AI. A generally accepted definition is McCarthy's (1956) who defines AI as the science and engineering of making intelligent machines. The following definitions of AI have been adopted by this research:

1 AI can be defined as some software that copies (or mimics) and produces human behaviours such as planning, the generation of ideas and being able to scale human intellect (Soral, 2023).

2 The good artificial intelligence is to create computers that can behave like humans and complete jobs that humans would normally do (Similar, 2023).

3 AI ethics is a system of moral principles and techniques intended to inform the development and responsible use of artificial intelligence technology (Lawton and Wigmore,

2023).

Next, a literature review is presented including the main research questions. This is followed by the Research Methodology, Data Collection and Interpretation, Results, Discussion and Conclusion sections. Appropriate and fit for intended purpose recommendations are presented last.

1.2 Literature Review

1.2.1 Humanistic Concerns

Duggal (2023) considers that artificial intelligence makes it possible for a computer programme to think and learn unaided. It simulates human intelligence within machines so that these machines can do things without human interferences. There are three types of AI: weak (one task capability limitation), strong (an ability to comprehend and learn any intellectual task (just like a human being can) and super (exceeding human intelligence, ability to perform any task superior to a human being). Table 1 summarises the advantages/disadvantages considered by Duggal.

Table 1. Advantages/Disadvantages of AI Applications, no order of priority, adapted from Duggal (2023)

Advantages	Disadvantages
Reduction in Human Errors (increase accuracy and precision)	High Costs (lots of resources needed to achieve and maintain)
Reduce risks by AI doing the job (such as defusing a bomb)	No creativity-AI cannot yet think outside any parameters
24/7 Availability versus the human being's productivity maximum of 3-4 hours a day)	Unemployment-has the potential to reduce human resources, thus leading to higher levels of employment
Applying digital assistants (reducing extensive human resources)	Has the potential to make human beings 'lazy'-reduction in using the human brain cells could seriously affect future human generations
Inventing Innovations across all fields	Unlike human beings, AI does not yet have ethics capability methods. Has the potential to wipe out humanity in future
Decisions based on logical and rational rather than emotions and feelings, leading to more accurate decisions	It appears that feelings and emotions, important for people, cannot be replaced by computers
More effective and efficient completion of repetitive tasks to free up human beings to focus on more challenging/creative tasks	At present, no improvements, AI can only do what it has been programmed to do
Best possible route planning to get to destinations	
Apply in risky situations such as flying to Mars or exploring the deepest oceans	

BBC News (2023) reports that AI has the potential to replace human jobs, carries some primary concerns about personal data issues such as privacy invasion and could be used to deceive/manipulate (false information, fake news). It appears that both Google and Microsoft, for example, developed similar ethical AI considerations and principles for use within their research and products. In addition, the Organisation for Economic Co-operation and

Development (OECD, 2019) developed some AI guiding principles that promote the innovative, trustworthy, and respectful use of AI to maintain human rights and democratic values.

Andersen and Rainie (2018) purport that it appears that there are opposing views as far as AI and its value to human beings are concerned. Some experts claim that AI will enhance people's lives, many others have raised concerns in areas such as being human, being productive and being able to exercise free will (based on the outcome of some canvassing of experts (979) during 2018. Major concerns included a threat to human autonomy and capabilities, and that computers might exceed human intelligence. On the positive side, they recognised that AI enables 'smart systems' in cars, buildings, and utilities. Health care, for example, would benefit enormously from AI applications. Many experts considered that they were concerned about AI's impact on what it means to be human. The consensus amongst the respondents from the 2018 research confirmed that human beings need to make sure that technology matches human values. Concerns raised by those interviewed include the following: 1 Loss of control over individual life 2 Data Management designed for profit or power gain 3 Social unrests due to job losses 4 Human beings cognitive, social and survival skills will degenerate 5 Increase in cybercrime and autonomous weapons. Potential suggested solutions to fix these problems include to improve global working together (including stakeholders), to develop policies that ensure humanness and common good and to put people first, robots second.

According to Wikipedia (2013), artificial general intelligence could ultimately lead to human extinction or some similar other unrecoverable global catastrophe. Wikipedia argues that the human specie's superiority over other species is driven by the distinctive capabilities of the human brain. Assuring and suggesting that AI has the potential to surpass these human brain capabilities, leading to AI becoming impossible to be controlled by human beings. It appears that two major concerns have been identified: control and alignment. Trying to control a super-human machine may be more difficult/impossible to achieve than first thought/considered. Any superior intelligence, by definition, would resist to be shut off or to change its goals. It would equally make it more difficult and potentially impossible to align this to human values and constraints. It is also possible that such a sudden intelligence explosion might be too much for humans. Technical improvements/capabilities will grow exponentially at a speed that may be impossible for human beings to compute and cope with. Progress from basic to beyond super-human may be too fast for human beings to cope with.

Stahl et al. (2023) report that the topic of ethical considerations associated with artificial intelligence (AI) appears to be a major challenge of the 21st century. Although AI carries many potential benefits such as operational improvements, these are sometimes outweighed by ethical concerns, including health and safety. A lack of detailed understanding of developments in AI can lead to incorrect determinations of ethical issues. Stahl et. al reviewed actual AI case studies and summarised their findings/conclusions. Ethical concerns were raised in the following areas (Table 2): discrimination, privacy, personal data, manipulation, right to life, dignity, and sustainable development goals (UN, 2015). Stahl et al. argue that AI systems should be programmed to avoid killing or harming human beings,

maintaining the fundamental human right to life, liberty, and security, in areas such as transport (self-driving cars), training/education (scoring of exams), law enforcement (reliability of evidence), employment (CV sorting in recruitment) and medical (robot-assisted surgery). AI developers clearly have a responsibility to minimise/avoid vulnerabilities and threats. Human life should be supported by AI and its effectiveness not lessened. The relevant and related concept of dignity in AI forms part of human rights (previous section). It is often considered to be of prime value by human rights supporters. According to Schroeder and Bani-Sadr (2017), dignity in AI ethical debates is often considered too ambiguous and difficult to assign.

It appears that discrimination is a major concern linked to AI (Latonero, 2018; Muller, 2020). This is based on observations of potential misappropriation of data leading to intended/unintended discrimination such as ethnic misrepresentation in recruitment. An AI System can establish past successful/unsuccessful applications, and thus make a prediction of future success. Predictive policies, using automated biometric recognition, provides police officers with past-history risk scores, potentially discriminating against people with previous convictions. A further potential issue is possible when an AI system is programmed to focus on data that does not fully represent the current population's data such as gender, age, or ethics. Privacy and data protection concerns, driven by advances in AI, have been raised: a.) Authoritarian regimes that employ AI would benefit from deeper AI data analysis, thus enabling appropriate Government interventions b.) The analysis of genetic data could lead to medical perceptions c.) AI appears to be a threat to privacy, adding new capabilities to cause personal harm to people. Another area of concern is the misappropriation of personal data for profit-making. Zuboff (2015) purports that information capitalism (domination of information and flow across the globe) aims to predict and change human behaviour as a way of producing revenue and exercising market control. This is an example of the illegal, unauthorised, or unfair adoption/collection of personal data without informed consent. The United Nations Educational, Scientific and Cultural Organization (UNESCO, 2021) suggests that:

- The AI method chosen should be appropriate and proportional to achieve a given legitimate aim.

- The AI method chosen should not infringe upon foundational values; its use must not violate or abuse human rights.

- The AI method should be appropriate to the context and should be based on rigorous scientific foundations.

Manipulation is yet another concern raised about the application of AI. This covers impacts at both personal and societal level. So-called data analytics techniques have the potential to be used in electronics, for example, to influence voter behaviour. Buying behaviour of consumers can equally be influenced through hetero-indoctrinated AI marketing ('AI-driven personal marketing', p. 55). Infringements include people's deprivation of making informed decisions. It is also possible to apply AI technology to give impressions that only certain options/choices are possible, thus violating autonomy.

According to Coplin (2019), AI should be designed to achieve humans plus machines, not humans versus machines. So-called care robots can provide some form of human dignity to elderly people, for example (also known as assistive robots). Monitoring robots can achieve this, too. In contrast, so-called companion robots appear to undermine elderly people's self-respect if offered to replace a face-to-face human interaction. This approach provides independence and dignity. When the word dignity is removed, all that remains is independence. Macklin (2003) suggests that AI in elderly cases maintains primarily their independence.

Zardiashvili and Fosch-Villaronga (2020) identify eight main ethical concerns related to the use of care robots for elderly people: safe human-robot interaction, the allocation of responsibility, privacy and data protection loss, autonomy restriction, deception and infantilising, objectification and loss of control, human interaction decrease and long-term consequences. The ethical concerns regarding dignity may ultimately be driven by the projected need for more social care workers, for example, in the United Kingdom, by 2035 (Macdonald, 2020). And finally, so-called sustainable development goals (SDG) were agreed by the United Nations (UN, 2015) to ensure that human beings are looked after as best as possible. For the SDGs to work, it will be necessary for all AI researchers to work closely together, by collaborating and engaging irrespective of resource-limited settings. It appears that is not necessarily the technological solutions that cause ethical AI concerns, but rather how AI is applied in practice and its reliability. Human vigilance and input are a constant requirement. Opacity of AI neural networks is essential for ethical reasoning. It will also be necessary to include the nature of an AI system explicitly in ethical reasoning (Stahl et al.).

Larson (2021) reasons that, based on the outcomes of some research, human and machine intelligence are fundamentally different. Low versions of intelligence appear to be applied, benefiting from improved /increased competing power, but not making accordingly appropriate progress. The unknown should be explored more, instead of focusing on promoting and pushing existing methods, using hyping-up tactics. Larson suggests that genuine inventions should not be pushed aside at the expense of some futuristic debates. Human beings should pay more attention to the one and only true intelligence that appears to exist-human beings' own. Consideration should be given to 'cheap imitations of deeper ideas that cut off intelligent engagement' (p.4). Interesting insight questions to ask should include whether it is possible to instil in the human mind's intuition capability (grasping truth and meaning) to a machine or computation. Human beings can see something that a machine/computation cannot, such as probability of truth (Goedel, 1931). Larson reports that some researchers argue that if a machine is fitted with a so-called learning system or learning capability to learn a particular task, then it is highly likely that the machine will perform poorly on other tasks. From an ethical perspective, it is worth noting that human beings, so far, have not been able to create more intelligent versions of themselves.

A related ethical consideration reveals that a gap exists between actual progress made and what the futuristic visions of computer scientists describe, such as Turing (1950), Good (1965), Vinge (1993) and Kurzweil (2005). This relates particularly to lack of progress being made in 'difficult aspects of natural language understanding' (p. 49, natural language includes

symbols, letters, punctuation and forming words, understanding context and the resolution of ambiguities in pronouns and indexical). The difference between human minds and machines appears to be greater than originally imagined by the mentioned scientists. The use of language is paramount and central to human being intelligence. This is a major concern and challenge for AI development, in both technical and humanitarian terms. Some of the identified and considered problems relating to natural language understanding could be addressed and improved, using statistical or machine learning approaches. Well-founded original issues raised by Haugeland (1989) remain valid in terms of semantics (meaning) and pragmatics (context). Russell (2019) argues that AI development needs to include elements of predicting human preferences. This could ensure that an AI learns more about what human beings really want and value. It would avoid AI systems from prioritising their objectives. In addition, AI should be developed to avoid any development of altruism towards human beings. Any learning should enhance performance, based on actual experience. Larson considers that better future AI performance can be achieved that ultimately leads to better and improved ethical performance, by focusing on problems in data-driven simulations and rather difficult ones that are not being analysed just based on frequency.

In their presented high-level overview of AI ethics, Kazim and Koshiyama (2021) report that the concept of AI ethics has developed, being driven by major concerns about AI impacts. It appears that there has been an increasing number of serious incidents of harm such as misuse of the technology (psychometric voter manipulation), facial recognition surveillance, mass data collection without consent or technological issues (bias in cases of recidivism). The authors' focus is on drawing attention to ethical issues and concerns across different disciplines, and to stimulate further thinking in the field. In this context, it is assumed that AI ethics is a sub-discipline of digital ethics. Machine-learning (ML) algorithms typically provide three types of learning: supervised (specified input to predict outputs), unsupervised (discovering unknown patterns) and reinforcement (trial and error-based decision made). Static algorithms (fixed sequence of actions) are not a major concern due to their fixed nature. Dynamic algorithms (learn and evolve through interaction) appear to cause concerns such as subconscious decision-making. Kazim and Koshiyama recommend that AI technologies need to be developed and deployed through appropriate laws. Enforcing legal compliance is an effective and positive approach to judge and escalate ethical AI issues and concerns. In addition, 'An ethical-by-design approach is a commitment to building systems ethically and in the hope that harm can be prevented' (p. 7). This includes the justification of design choices and accessibility to the system. Auditing and impact assessments allow for verification and accountability checks.

Stahl et al. (2022) report that progress has been made to address human rights implications of AI in the form of legal and regulatory proposals. The creation of a so-called 'AI Agency' is considered, with an unambiguous task to supervise all matters AI. Initial responsibility could be limited to Europe but global roll-out is potentially possible, too. This would ensure that any raised human and ethical rights issues can be addressed appropriately, including AI development, deployment, and use. The many economic and social benefits of AI need to be balanced with the raised concerns about ethical and human rights issues. There appears to be

a need for AI regulation, providing guidance, following on from responsible behaviour and collecting good practice. A so-called legislative framework needs to be developed and created, in the form of a responsible agency. Typical actions should include periodic updating of the agency itself to ensure continuous relevancy, managing concerns raised about ethical issues and human rights infringement, managing data protection issues, and addressing unfair discriminations. Other areas include but are not limited to employment, politics and AI systems becoming self-conscious.

Stahl et al. suggest that any legal interventions should not be viewed as being disruptive or interfering but rather as being supportive and enhancing to reach social goals and human thriving. This would bring together various AI understanding (technical, legal, and ethical) and could ultimately lead to the creation of Government-backed guarantees similar to kite marks), addressing human rights issues and concerns. On the negative side, the introduction of such a new regulatory body or agency could create new challenges, including overregulation concerns, regulatory capture (manipulation issues), distraction concerns (focus on risks rather than benefits) and introducing conflict (relating to AI-sensitive information such as Intellectual Property (IP) rights. Stahl et al. recommend that more research needs to be conducted to throw more light on how progress made in the considered areas could be measure reliably and validly.

The British Broadcasting Corporation (BBC, 2023) in the United Kingdom published a series of AI related articles. According to OpenAI (2023), an uncontrolled artificial general intelligence system has the potential to cause serious harm to the world. This applies equally to autocratic countries with a decisive superintelligence system. OpenAI suggests that AI needs to slow down at critical junctures. The United Kingdom Government appears to have produced plans to regulate AI to protect people's privacy, human rights, and safety. In contrast, some AI supporters claim that the new technology is delivering social and economic benefits for people. Too many regulations could jeopardise the progress that has already been made. Existing regulators in the United Kingdom such as the Health and Safety Executive (HSE) and the Equality and Human Rights Commission (EHRC) are likely to use existing laws to control AI regulation, focusing on five principles:

- Safety security and robustness

- Transparency and Ability to Explain'

- Fairness

- Accountability

- Contestability and Redress

The Investment Bank Goldman Sachs suggests that GAI can generate content that cannot be distinguished from work completed by human beings. Suggested is that AI's impact across different sectors will vary, from 46% of tasks in administrative professions to 6% in construction. Torsten Bell, Chief Executive Officer (CEO) of the Resolution Foundation, suggests that there is no clear evidence that suggests how the new technology will evolve or

how organisations will integrate the new technology into their 'modus operandi'. Torsten Bell considers that people should also focus on the positive contributions of AI, namely gains in living standards, higher productivity and cheaper to run services. According to the Future of Life Institute in the United States (2023), 'AI systems with human-competitive intelligence can pose profound risks to society and humanity' (letter signed by key figures in artificial intelligence, including Elon Musk (Twitter CEO), Steve Wozniak (Apple Co-founder) and some researchers at Deep Mind (2023).

1.2.2 Regulatory Positions

Andersen (2021) considers that AI ethics need to be taken to the next level: AI policy. An AI policy needs to be developed before AI gets out of control. Areas such as human rights and individual liberties should be on the top of the agenda. It appears that making policy is important to everyday life. The reason for this it that it contains universal human rights (Declaration of Independence in the United States, for example) and the ideals and goals of an entire nation. Whilst AI is already helping humankind in many areas including the automotive industry and medicine, it also poses new risks to human rights relating to discrimination, surveillance, transparency, privacy, and security, to name but a few. Ethical AI concerns must be addressed and corrected. This includes transparency of AI algorithms, bias/fairness of AI, algorithms resulting from the data/AI model training and risk assessment, and an appropriate approach to uncover any ethical issues and manage these. More policy frameworks (such as the UK Guidelines for AI Procurement or the Danish Principles for AI), need to be developed, controlled, and implemented. In addition, Andersen suggests that the following policy frameworks should be considered: 1 Mandatory Requirement of an AI Certificate: a kind of pre-qualified AI supplier list of AI products, services and operations 2 Obliged Risk Assessment: a Governmental risk assessment tool applied to supervise, control and mitigate any issues that relate to the roll-out of AI (including its data) 3 Dash-board Reports produced from the commercial data's economic value of the end user plus details of exactly what data has been collected 4 Audited AI Accountability Report: any AI supplier/deployer must show evidence of ethical and human rights concerns regarding AI, and how they intend to deal with any concerns.

Globallegalpost (2023) reports that Denmark was planning to introduce mandatory company legislation for AI and data ethics on 1 July 2020, for compliance by 2021. As of 23 November 2022, it appears that specific regulations for AI are still not in place. AI is covered by current data protection law, legal principles apply. In the UK, there are no specifically written AI laws in place. Some elements of AI are regulated through a series of existing legal and regulatory requirements. This includes and covers AI technologies. New proposals are progressing to regulate the use of AI technologies. Germany is still in the process of developing a specific AI law. The aim of the new law is:

- To become a leading AI location to secure future competitiveness.

- To ensure responsible and public welfare-oriented development and use of AI.

- To ensure that AI applications are ethical, legal, cultural, and institutional, aligned to

existing social co-existence principles.

Affected areas include fundamental human rights, bias and discrimination, intellectual property, patents, copyright, trade secrets/confidentiality and data protection, including biometric data (voice and face).

Alston and Bird (2022) and Felz (2022) report that some new AI regulations are planned for roll-out in the US during 2023. This includes industries such as financial services, insurance, automotive, logistics, health care and medical devices, retail e-commerce and advertising and marketing. Expected AI regulations for 2023 include state data privacy, an AI risk management framework, Federal Trade Commission (FTC) rulemaking and new National Institute for Standards and Technology (NIST) AI standards.

1.2.3 Specific Purpose Applications

Herath and Mittal (2022) purport that AI has been used in smart city research for over 15 years. Urbanisation has had a major impact on the environment, in associated areas such as management, healthcare, energy and education. It appears that the concept of AI has been applied in so-called 'smart city' developments. Applying AI in smart city solutions has distinct advantages in areas such as improved adequate water supply, energy and waste management and reduced traffic noise and pollution. Although smart city solutions provide benefits such as efficiency improvements and automation, they also generate some regulatory issues, including service delivery and privacy/ethical concerns. According to Tecuci (2012), AI is used to train computers to imitate thinking processes and human behaviour. The driving forces and impact areas behind AI's increasing involvement in smart cities are shown at Figure 1.
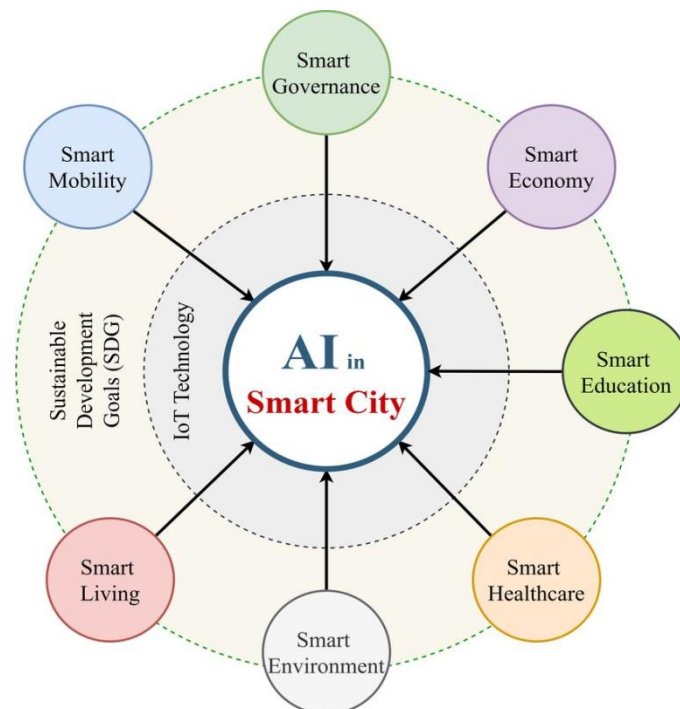

Figure 1. The Driving Forces and Impact Areas of AI in Smart Cities (ResearchGate, 2023)

The advantages of using AI to develop smart city concepts appear to outweigh any adverse connotations. Domains such as smart energy, mobility/transportation, health care and agriculture are already playing an important part in every-day life. In smart energy, AI has the potential to produce energy forecasting that human beings could not produce, including the training of an energy consumption prediction model. In transportation, AI could be applied to enhance so-called intelligent transportation systems (ITS) that will enhance transportation in smart cities, substantially. Another example is agriculture where AI is already engaged in automated techniques to optimise the use of labour, manage climate change impacts, and improve food security/population increase challenges. In contrast, these applications of AI across different domains carry testing regulatory disputes and challenges, including data availability, discrimination and privacy, and a shortfall of qualified AI professionals.

Liwang (2022) claims that the application of AI within the military (defence environment) in Sweden carries some societal challenges such as the ones raised by the European Commission (2020). Concerns include effects on national security and a lack of ethical values and principles embedded in the developed technology. Liwang reports that that there are further societal challenges including guiding principles of AI development, AI training and decision-points how and when AI should be applied. It appears that current AI policy does not represent or address such concerns.

Terkonda and Fish (2023) consider that, in the United States of America, for example, medical boards appear to have been driven to respond to major regulatory framework challenges that relate to the use of artificial intelligence in medical care. The area of social responsibility is critical in the ever-growing adoption of AI in health care. It appears that a new innovation-friendly regulatory strategy is required that will allow AI to be realised whilst at the same time improve the safe health and welfare of patients. State medical boards established considerate levels of trust over the last 150 years since their inception within the community of health care practitioners. Their primary objective is to ensure that safe care is established and maintained, including the application of AI in clinical settings under relevant regulatory supervision and associated accountability. Putting these controls in place provides a framework for the definition of the role of AI in medicine. Specific emphasis is placed on humanistic values such as respect, integrity and providing patients with guidance by adopting a shared decision-making process. According to Hinton (2016), it may be possible for AI to act as a proxy, for example, in place of a physician in radiology. Terkonda and Fish argue that physicians can be expected to obtain permissions from patients as far as the collection and use of patient data is concerned. Any algorithm-based statement needs to incorporate a flexible statistical analysis ability that allows for culturally different data to be processed appropriately, thus avoiding risks of harmful, inaccurate, or misleading results. In addition, different regulators (by sector, by country) need to support each other to try and achieve some regulatory goals at the same time. This should be a highly desirable target set by the involved international and regional regulatory entities. A so-called 'systems thinking' approach (including patient involvement) has the potential to become a pro-active approach to safeguard human beings' ethical values within such a collaborative environment.

Roberson et al. (2022) report that Australia's Chief Defence Scientist states that AI

technology must be given ethical and technological development considerations, in parallel with the benefits such technology offers, quoting Ziesing, 2021). Roberson et al. consider that an approach known as 'responsible research and innovation' (RRI) should be applied to achieve ethical design approaches and principles during the whole development life cycle of technologies such as AI. RRI approaches incorporate societal and technical challenges driven by AI. It includes establishing ways and means of developing ethical and safe human-AI partnerships (Ramchurn et al., 2021), clearly defining the responsibilities of both human and AI agents (Yazdanpahah et al., 2012), and reviewing/investigating the ability to explain methods for autonomous systems (Omeiza et al., 2021).

Robertson et al. suggest that such control systems will help organisations to introduce changes responsibly, and within military environments to help commanders to lessen ethical risks. In addition, Article 36 (Geneva Convention, 1949) appears to be followed by international military organisations to comply legally and to ensure humanitarian and ethical compliance, too, when it comes to the development and future applications of AI systems and software. Such a system, known as Athena AI, has been designed and developed in such a way that it identifies objects and people who cannot be targeted on the battlefield. Of equal importance is that Australia's Department of Defence follows ethics-by-design principles (D'Aquin et al., 2018; Dignum, 2018; Shilton, 2013), in line with a published report 'A Method for Ethical AI in Defence (MEIAD, Devitt et al., 2021). This report presents five domains for establishing ethical AI: 1. Responsibility (who is responsible?) 2 Governance (how is AI controlled?) 3 Trust (how can AI be trusted?) 4 Law (how can AI be used lawfully?) 5 Traceability (how are the actions of AI recorded?)

There are two more AI safeguards in place in Australia: 1 Data Ethics Canvas: data collection, ability to trace data back to its original source 2 Ethical AI risk matrix: tracking of ethical risks, highlighted risk reductions/mitigation, owners of these risks, plus when the mitigations will occur.

Table 2 depicts a summary of the main identified ethical challenges.

Table 2. Ethical Challenges Summary, adapted from Stahl et al. (2023)

Topic	Impact/Area	Evident Effect	Concerns	Mitigation	Conclusions
Discrimination	Age, Race, Gender, Disability	Gender Bias, Predictive Policing, Skin Colour	Machine-learning system, Unfair Discrimination	Impact Assessment, Ethics by Design	AI systems could make it worse but have potential to identify it.
Privacy	Private Data, Genetic Data, Biometric Data	Violation of Data Protection, Moral Values	Profiling, Honest Dealings, Mission Creep	Data Protection Impact Assessment System of AI Application	Recognised as Human Right Legislation
Personal Data	Inequality, Ruling by Data	Breach of Data Protection Regulation, Individual Rights	Segmenting, Deception, Replacing Beliefs	Antitrust Regulation, Data Sharing and Access	Whistleblowing, Policy, and Societal Interventions
Manipulation	Political, Vulnerable Consumers	Data Analytics Techniques, Hetero-Indoctrinated AI Marketing	Ethical Evaluations, Transparency of Data	Standardisation, Ethics-by-Design Methodology	Related to other ethical concerns, No Single Solution in Place
Right To Life	Transport, Home Security, Healthcare, Law Enforcement, Employment	Self-Driving, Exams, Managing People, Credit Scoring	Security and Privacy, Deep Learning, Human Safety	Effective Liability Regime, Quality Management System	AI programmed not to kill/injure Humans, must support Human Life, not Undermine it
Dignity	Human Rights, Automation, Care Robots	Independence, Self-Respect	Thinking AI is infallible, Infantilising	Use all Ethical Values	Dignity Considerations, Override other Ethical Values
Sustainable Development Goals (SDG)	Resource Limited Settings, Sustainability	Precision Agriculture and Farming, Data Input	Lack of AI Training, Lack of Partnerships	Closer collaboration of resource-limited resources	Closer collaboration between AIs and Information and Communication Technologies (ICTs)

1.3 Film Review

M3GAN (2023) is an artificial but life-like doll, designed to be the companion of any child, supporting the parents. The doll can listen, watch, and learn, being listener, watcher, and learner at the same time. The prototype doll develops some unplanned and unforeseen attitudes and behaviours that lead to disastrous consequences (through the development of self-awareness). M3GAN stands for Model 3 Generative Android, capable to produce and reproduce). It appears that M3GAN is left to her own devices, basically disregarding her own autonomy, and asking herself at regular intervals to shut down. This has a major impact on

her development. She is left on her own to acquire new knowledge and experience and then find her own way of interpreting it. So-called 'machine-learning' is currently debated heavily across industries, according to Ash (2023).

The production of this toy known as M3GAN appears to be commercially driven to become the toy that will replace all other toys. Capabilities include to become autonomous, smell sensors, emotional responses and repetitive commands that ultimately develop emergent capabilities. Some programming issues that developed during M3GAN's early existence include human interaction capability such as protecting the human child whilst showing an increasing disregard for other people, a misguided attachment theory (in lieu of a real person) and appropriate social skills around other people. The built-in concept of 'learn, calibrate and optimise' appears to have failed together with the doll's claimed palliative capability to self-heal including contextual relevance such as managing and controlling distressing symptoms.

1.4 Main Research Questions

The main research questions for this research are:

1. What is known about the development and application of AI in work environments?
2. Is there a relationship between increases in AI business applications and adverse effects on human beings?
3. Does AI maintain sufficient respect for the avoidance of privacy invasion and the rights of individuals?
4. Should ethical considerations outweigh advantages of AI applications across industries?
5. Has the introduction of AI at work created any management issues and concerns?
6. Should AI become capable of self-development, for example, in terms of social skills, in artificial intelligence environments?

2. Research Methodology

2.1 Method

The researcher adopted a structured qualitative research approach to capture the essence of contemporary thinking, knowledge, and experience about the topic under scrutiny. 6 central attention of the literature review was on publications between 2020 and 2023, with occasional references dating back to the 1900s. This approach enabled the answering of the research questions within the chosen research field. The aim was to organise, analyse, and interpret both the theoretical and practical application impacts of AI at work and on people, using non-numeric and conceptual information. The research scope was divided into three domains, namely: humanistic concerns, regulatory positions, and specific purpose applications. It was thus possible to identify patterns and themes within the captured data (thematic analysis, Section 2.2). The results produced new thinking and previously unknown approaches. In addition, the researcher completed a critical and systematic review of the identified and reviewed theories, with particular emphasis on their validity and accuracy. The research framework was narrow. It investigated the relationship between the chosen three topics of

applied AI. It produced a solid overview of the identified current research problem (Introduction). The research's primary focus was on the development of theory and practice in AI applications.

2.2 Data Collection and Interpretation

The researcher identified appropriate research data from the literature review through a quasi-thematic analysis approach (the amount of generated data did not warrant the use of coding). Research findings focused on evidence-based analysis and interpretation of the congregated data. All data was analysed by theme/topic to assign meaning to the data to arrive at relevant conclusions (quasi-thematic analysis). The adopted data analysis and interpretation included data collection, developing findings, and developing conclusions and recommendations. In addition, predictive analysis was adopted to suggest potential and likely future AI approaches and applications ('what is likely to happen scenario'). Although this generated some conjectural knowledge within unknown areas, it contributed to the development of accurate predictions. An overall summary of the identified and reported ethical concerns across different domains is shown at Table 3. The identified and ranked as critical essential ethical examinations, by domain, are presented at Table 4.

Table 3. Summary of Ethical Considerations/Concerns by Research Domain

Humanistic Concerns	Regulatory Positions	Specific Purpose Applications
Duggal (2023): Increase in unemployment, reduction in human brain cells, destroy humanity, no feelings/emotions.	Globallegalpost (2023): mandatory AI company missing.	Herath & Mittal (2022): optimise use of labour, manage climate change, improve food security/population increase.
Pasanen (2022): privacy invasion, misuse of personal data, higher unemployment.	Andersen (2021): AI policy missing, mandatory AI certificate needed, audited AI accountability needed.	Terkonda & Fish (2023): healthcare sector, adopt a shared decision-making process in healthcare, reduces risk of harmful, inaccurate, or misleading results.
Andersen & Rainie (2018): threat to human autonomy and capabilities.	Alston & Bird (2022): AI regulations not in place, data privacy and risk management framework missing, AI standards needed.	Robertson et al. (2022): military applications of AI to reduce ethical risks by developing ethical and safe human-AI partnerships.
Wikipedia (2013):AI cannot be shut off, human extinction, AI development too fast for humans.	Stahl et al. (2023): legal and regulatory proposals need to be developed; responsible agency needed (legislative framework), Government-backed guarantees required, intellectual property (IP) rights need addressing.	
Stahl et al. (2023): health and safety issues, discrimination, personal harm, misuse of personal data, manipulation, infringements, more AI collaboration of those who work in AI.		
Larson (2021): ability to grasp truth and meaning missing, reduction in altruism towards human beings, misuse of technology, lack of consent for data collection, bias in cases of recidivism.		
BBC (2023): plans appear to have been produced by the UK Government to protect privacy and human rights.		
Universal Pictures (M3GAN, 2023): human competitive intelligence poses serious risks to humanity, increasing disregard for other people, misguided attachment theory, missing is the ability to manage and control distressing symptoms, disregard for other people, self-healing contextual relevance.		

Table 4. Essential Ethical and Considered Critical AI Examinations

Ethical AI Considerations (positive and negative)	Ranking (0=Not important, 1=Neutral, 2=Important, 3=Very Important)
1. Humanistic:	
-Increase in Unemployment	3
-Misuse of Personal Data	3
-Destroy Humanity	2
-AI Development too fast for Humans	2
2. Regulatory:	
-Legislative AI Framework Needed	3
-AI Agency Required	3
-AI Accountability needs to be Audited	3
3. Specific Purpose Applications:	
-Optimise use of Labour	2
-Reduce risks of inappropriate Results	2
-Reduce ethical risks in the Military	3

3. Results

The outcomes from this research established that sufficiently great progress has been made in AI to warrant taking a fresh look at potential associated ethical issues at humanistic, regulatory, and special purpose application levels. Gaining an understanding of in-depth knowledge and recent developments in AI capabilities has led to the revelation of technical solutions that could lead to either gigantic improvements to the modus vivendi of human beings or their ultimate demise/extinction. Progress in AI appears to have reached a critical point in its current development. Two contrasting and opposing views suggest that AI applications have both positive and negative implications for human beings. On the positive side, some experts consider that AI will improve people's lives, for example, in areas such as healthcare, the service industry and smart systems in cars, buildings and utilities. On the negative side, major concerns have been raised about making more and more people unemployed, a reduction in human autonomy and capabilities and that, ultimately, computer systems could exceed human intelligence and become the catalyst for human extinction. A recently released thought-provoking film (M3GAN) presents further insights into what could happen if so-called 'machine learning' capabilities are introduced that show, for example, no regard for emotional attachment, social skills and managing/controlling distressing symptoms. It appears that the identified benefits for human beings, according to and reviewed by this research's selected authors, and business, are in serious disagreement with the major ethical concerns raised by the same authors. There is an increasing pressure on Governments across the world to develop and implement effective and appropriate legislations/legal frameworks to protect the best interests of human beings/societies. This includes and is not limited to human rights, individual liberties, and discrimination. Humanistic concerns appear to be of paramount importance, followed by regulatory concerns and considerations. Some countries have been considering and planning to roll out mandatory legislation such as Denmark and the United States. This research confirmed that these planned legislations have not been rolled out yet at the time of completing this research paper.

4. Discussion

A relationship appears to emanate from this primarily contemporary research that suggests there is a correlational increase in ethical AI concerns driven by the rapid development of AI capabilities. The AI pioneering work of the past (McCarthy, 1955, was the first person to be credited with the use of the term Artificial Intelligence) created some valid and reliable application solutions that appeared to be 'safe' for human beings. This changed dramatically over the last five years. Recent application developments make it possible, for example, to reduce human errors, reduce loss of human life and enable 24/7 non-stop working. Job losses and higher levels of unemployment could be the potential consequences, affecting human communities exponentially. In addition, major concerns have been raised such as machines becoming more intelligent than humans, thus negating the need for any human resource/presence. If this development is allowed to continue unrestricted, there is a high risk that machines could ultimately erase humankind. Privacy invasion, misuse of personal data and discrimination issues to name but a few, have become major confidentiality breach worries. If left unchanged, this could lead to AI systems being allowed to operate feely without feeling or emotional responses. Advanced emotional attachment AI software may develop too strongly and become difficult to control/untangle (M3GAN). In contrast, there are some positive AI characteristics to report. The application of AI in healthcare has led to marked improvements such as more appropriate medical decision-making, reducing the risk of harmful, inaccurate, or misleading diagnostics. AI is also used to improve the process of managing climate change or improving the security/safety of food. In the military, AI is applied to reduce ethical risks such as innocent bystanders being harmed by military intervention. Irrespective of whether people feel positive or negative about the further development of AI, there appears to be an underlying trend by both parties that implies that more needs to be done to develop and introduce appropriate clearly defined AI policies and regulations. Specific suggestions include the need to introduce some data privacy and risk management framework, audited AI accountability and AI standards.

5. Conclusions

AI has the potential to become the new de facto standard for the simulation of human intelligence processes by machines, particularly computing systems. Identified associated ethical concerns, to make this work in the best interests of humankind, need to be taken into consideration when developing, testing, and rolling out AI systems and related software applications. In addition, more appropriate control systems need to be introduced to safeguard against the growing number of concerns/issues that have been raised in recent times, ranging from humanistic to regulatory and specific purpose worries. The introduction of pertinent legislation together with relevant mandatory AI control systems such as risk assessment, an AI pre-qualified supplier list of products, services and operations, and audited AI accountability reports (evidence of ethical and human rights concerns, and how they intend to manage these), would be a good starting point to exercise reasonable and realistic jurisdiction. More international collaboration and co-operation will be necessary to ensure that the identified AI ethical concerns can and will be addressed across boundaries for the benefit of citizens world-wide, irrespective of their geographical location and cultural diversity. More

research needs to be conducted at AI practitioner level. The insights of those who are, and have been working in this area, would make a welcoming addition to the AI body of knowledge. A series of face-to-face interviews together with an associated focus group meeting would allow the close bringing together of theory and practice. The researcher confirms that all research questions (1.3) have been answered by this research.

References

Alston and Bird. (2023). [Online] Available: https://alston.com

Andersen, J., & Rainie, L. (2018). *Artificial Intelligence and the Future of Humans*. PEW Research Center. [Online] Available: https://www.pewresearch.org

Andersen, K. S. (2021). *Moving From AI Ethics to AI Policy*. [Online] Available: https://2021.ai

Ash, E. (2023). *'M3GAN and the Ethics of the AI Toy Shop*. Willamette Week. [Online] Available: https://www.wweek.com/movies

BBC News. (2023). *AI could replace equivalent of 300 million jobs-report*. [Online] Available: https://www.bbc.com

BBC. (2023). [Online] Available: https://www.bbc.com

Coplin, D. (2019). *This is the rise of the humans! Dispelling the myth about the robot uprising*. [Online] Available: https://www.fenews.co.uk

D'Aquin, M., Troullinou, P., O'Connor, N. E., Cullen, A., Faller, G., & Holden, I. (2018). *Towards an 'ethics by design' methodology for AI research projects*. In Proceedings of the 2018 AAAI/ACM Conference on AI, ethics and society. pp. 54-59. https://doi.org/10.1145/3278721.3278765

Deep Mind. (2023). [Online] Available: https://www.deepmind.com

Devitt, K., Gan, M., Scholz, J., & Bolia, R. (2021). *A method for ethical AI in defence*. [Online] Available: https://parlinfo.aph.gov.au

Digital Strategy. (2023). [Online] Available: https://digital-strategy.ec.europa.eu

Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Information Technology, 20*(1), 1-3. https://doi.org/10.1007/s10676-018-9450-z

Duggal, N. (2023). [Online] Available: https: //www.simplilearn.com

European Commission. (2020). *A European Approach to Excellence and Trust*, White Paper on Artificial Intelligence, European Commission, Brussels.

Felz, D. J. (2022). [Online] Available: https://www. alston.com

Future of Life. (2023). [Online] Available: https:// www.futureoflife.org

Geneva Convention. (1949). [Online] Available: https://en.m.wikipedia.org

Globallegalpost. (2023). [Online] Available: https://www. globallegalpost.com

Goedel, K. (1931). (translated version by Meltzer, B., 1992). *On Formally Undecidable Propositions of Principia Methematica and Related Systems*, Dover Publications, Inc., New York

Good, J. I. (1965). Speculations Concerning the First Ultraintelligent Machine. *Advances in Computers, 6*, 31-88. https://doi.org/10.1016/S0065-2458(08)60418-0

Haugeland, J. (1989). *Artificial Intelligence-The Very Idea*. Cambridge, MA MIT Press. https://doi.org/10.7551/mitpress/1170.001.0001

Herath, H. M. K. K. M. B., & Mittal, M. (2022). *Adoption of artificial intelligence in smart cities: A comprehensive review*. *Int Journal of Information Management Data Insights, 2*, 100076. https://doi.org/10.1016/j.jjimei.2022.100076

Hinton, G. (2016). *Creative destruction lab-on radiology*. [Online] Available: ttps://www.youtube.com/watch?y1/421MPRXstSvQ

Kazim, E., & Koshiyama, A. S. (2021). *A high-kevel overview of AI ethics*. Patterns, Cell Press. https://doi.org/10.2139/ssrn.3609292

Kurzweil, R. (2005). *The Singularity is Near: When Humans Transcend Biology*. New York Penguin Group.

Larson, E. J. (2021). *The Myth Of Artificial Intelligence…Why Computers Can't Think The Way We Do*. Harvard University Press. https://doi.org/10.4159/9780674259935

Latonero, M. (2018). *Governing artificial intelligence: upholding human rights & dignity, Data & Society*. [Online] Available: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf

Lawton, G., & Wigmore, I. (2023). *AI ethics (AI code of ethics*. [Online] Available: https://techtarget.com

Liao, S. M. (2020). In '*A Short Introduction to the Ethics of Artificial Intelligence' (Ethics of Artificial Intelligence* (Ed.). Oxford University Press. https://doi.org/10.1093/oso/9780190905033.001.0001

Liwang, H. (2022). Defense development: The role of co-creation in filling the gap between policy-makers technology development. *Technology in Society, 68*, 101913. https://doi.org/10.1016/j.techsoc.2022.101913

M3GAN. (2023). *Universal Pictures*. [Online] Available: https://www.universalpictures.co.uk

Macdonald, M. (2020). *The health and social care workforce gap*. Insight, 10 January, House of Commons Library, London. [Online] Available: https://commonslibrary.parliament.uk/the-health-and-social-care-workforce-gap

Macklin, R. (2003). Dignity is a useless concept. *BMI, 327*, 1419-1420. https://doi.org/10.1136/bmj.327.7429.1419

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. [Online] Available: https://raysolomonoff.com/dartmouth/boxa/dart564props.pdf

Micology. (2023). [Online] Available: https://micology.com

Muller, C. (2020). *The impact of artificial intelligence on human rights, democracy, and the rule of law*. Ad Hoc Committee on Artificial Intelligence (CAHAI), Council of Europe, Strasbourg. [Online] Available:
https://rm.coe.int/cahai-2020-06-fin-c-muller-the-impact-of-ai-on-human-righs-democracy-/16809ed6da

OECD. (2019). [Online] Available: https:// www.oecd.org

Omeiza, D., Webb, H., Jirotka, M., & Kunze, L. (2021). Explanations in autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 1-21. https://doi.org/10.1109/TTTS.2021.3122865

Ramchurn, S. D., Stein, S., & Jennings, N. R. (2021). Trustworthy human AI partnerships. *Science, 24*(8), 1-13, https://doi.org/10.1016/J.ISCI.2021.102891

ResearchGate. (2023). [Online] Available: https://www.researchgate.net

Robertson, T., Bornstein, S., Liivoya, R., Ng, S., Scholz, J., & Devitt, K. (2022). A method for ethical AI in defence: A case study on developing trustworthy autonomous systems. *Journal of Responsible Technology, 11*, 100036. https://doi.org/10.1016/j.jrt.2022.100036

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking, New York.

Schroeder, D., & Bani-Sadr, A-H. (2017). *Dignity in the 21st century: middle east and west*. Springer Int Publishing AG, Cham. https://doi.org/10.1007/978-3-319-58020-3

Shilton, K. (2013). *Values levers: Building ethics into design. Science, Technology and Human Values, 38*(3), 374-397. https://doi.org/10.1177/0162243912436985

Simplilearn. (2023). [Online] Available: https://www.simplilearn.com

Soral, S. (2023). [Online] Available: https://www2.deloitte.com

Stahl, B. C., Rodrigues, R., Santiago, N., & Macnish, K. (2022). A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. *Computer Law & Security Review, 105661*. https://doi.org/10.1016/j.clsr.2022.105661

Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). *Ethics of Artificial Intelligence-Case Studies and Options for Addressing Ethical Challenges*, *Springer.* https://doi.org/10.1007/978-3-031-17040-9

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, 59*(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

United Nations. (2015). *Transforming our world: the 2030 Agenda for Sustainable Development*. Resolution adopted by the General Assembly. [Online] Available: https://documents-dds-ny.org/doc/UNDOC/GEN/N15/291/89/PDF/N1529189.pdf?OpenElement

United Nations. (2023). *Department of Economic and Social Affairs*. [Online] Available: https://sdgs.un.org

Vinge, V. (1993). *The Coming Technological Singularity: How to Survive in the Post-Human Era of Cyberspace*. In G. A. Landis (Ed.), *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace* (pp. 11-22). NASA Publication CP-10129.

Wikepedia. (2013). *Existential risk from artificial general intelligence*. [Online] Available: https://www.wikepedia.org

Yazdanpahah, V, Gerding, E., Stein, S., Dastani, M., Jonker, C. M., & Norman, T. (2021). *Responsibility research for trustworthy autonomous systems*. In 20th International conference on autonomous agents and multiagent systems. pp. 57-62.

Zardiashvili, L., & Fosch-Villaronga, E. (2020). "Oh, dignity too?" said the robot: human dignity as the basis for the governance of robotics. *Minds, 30*, 121-143. https://doi.org/10.1007/s11023-019-09514-6

Ziesing, K. (2021). *Ethical AI report released from DST group*. Australian Defence Magazine, 1. [Online] Available: https://www.australiandefence.com.au/defence/cyberspace/ethical-ai-report-released-from-dst-group

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology, 30*(1), 75-89. https://doi.org/10.1057/jit.2015.5