



UWS Academic Portal

Artificial intelligence ethics

Simões, Gisele Barreto

Accepted/In press: 12/08/2020

Document Version
Peer reviewed version

[Link to publication on the UWS Academic Portal](#)

Citation for published version (APA):

Simões, G. B. (Accepted/In press). *Artificial intelligence ethics: a key factor in the development of mature information societies*. Paper presented at British Academy of Management Doctoral Symposium 2020, Lancaster, United Kingdom.

General rights

Copyright and moral rights for the publications made accessible in the UWS Academic Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact pure@uws.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Artificial Intelligence Ethics: a Key Factor in the Development of Mature Information Societies

Gisele Barreto Simoes
University of the West of Scotland
High Street, Paisley/UK - PA1 2BE
gisele.simoes@uws.ac.uk
+44 7485038103

Sustainable and Responsible Business

artificial intelligence; ethics; information society; regulation; sustainability

A qualitative research with focus groups and interviews.

I would appreciate help with the theory and methodology.

Artificial Intelligence Ethics: a Key Factor in the Development of Mature Information Societies

The paper presents the research purpose regarding the impact of artificial intelligence ethics in the development of mature information societies, and its emerging issues and regulation. This preliminary study examines through a literature review the main concepts of artificial intelligence, AI ethics, information society and the current stage of AI development, ethics issues and regulation. A primary qualitative research is suggested, with focus groups and interviews. The focus groups address the general public to understand the people's AI concerns, while the interviews focus to gain a deeper meaning of AI ethics from experts. The paper concludes with five research questions.

KEYWORDS:

Artificial intelligence; ethics; information society; regulation; sustainability

TABLE OF CONTENTS

RESEARCH PURPOSE	4
LITERATURE	5
METHODOLOGY	10
RESEARCH QUESTIONS	12
REFERENCES	14

RESEARCH PURPOSE

The research aims to investigate the impact of artificial intelligence in the mature information societies, exploring the AI information process system and its correlation with human intelligence, psychology and socialization, the new ethical issues emerging from its increasing integration in society and investigating the most appropriate frameworks to evaluate and regulate the impact of AI in the development of mature information societies from a moral perspective.

LITERATURE

Artificial intelligence or AI became again a trending topic in the 10s, with all sizes of companies inappropriately using the term for marketing purposes (Shead, 2020). Heikkinen (2019, p. 15) states the concept of artificial intelligence is hard to define but could be described from a legal perspective as

“a non-biological autonomous entity which has the ability to give rules to itself and the ability to make choices by an evaluative process”.

In the computer science field, AI is an area that investigates how computers can imitate the intelligence of human beings, with the ability to make judgements and decisions based on patterns (Sarmah, 2019). A concept extensively explored in the past by Weizenbaum (1976), who states working on AI is to build a machine on the model of the man, to learn language as a child does, and ultimately contemplates the whole domain of human thought. Another classic definition provided by Taddeo and Floridi (2018, p. 751), and recommended to address ethical and policy-related issues, describes AI as

“a growing resources of interactive, autonomous, self-learning agency, which enables computational artifacts to perform tasks that otherwise would require human intelligence to be executed successfully”.

AI ethical concern is not a new topic as well. Weizenbaum (1976) questioned the appropriateness to delegate all human functions to a machine, arguing computers and men are not the same kind of species. Similar view was held by Dreyfus et al (1986), who said machines are not in this world - they do not have body, childhood or culture practice - so they cannot acquire intelligence as a human being does. Other authors say machines cannot be intelligent as human beings since human intelligence is not algorithmic (Penrose, 1994), but organic; and as organic it is not measurable (Lebow et al, 2016). Regardless of decades of research in psychology and philosophy fields, there is no agreement about the definition of human intelligence (Lanz, 2000). However, the assumption that machine intelligence is identical to human intelligence is not mandatory to create intelligent machines; still, many AI researchers keep persuading the human-like intelligence goal (Fjelland, 2020).

Trend (2001) argues the digital era and the cyber-enthusiasm represented a radical break for society, raising the logic of mathematics over human irrationality. About cyber-enthusiasm, Linden and Feen (2003, p. 11) define the hype cycle as

“an useful educational tool for helping enterprises understand the inevitable pattern of excitement and disillusionment about technologies”.

Lebow et al (2016) remember Lewis Mumford’s vision, who said that the machines which are not built for human purposes put human individuality, principles and democracy at risk, making humans feel detached, passive and powerless. For Rifkin (2014), the technological revolution changed the way companies interact in the market, in the creation of new products and business models, an accelerated and complex process of development that Kotler (2015)

suggested could put corporations under pressure for short-term results. The operating system accessibility created by Richard Stallman and the Free Software Movement (Hughes, 2016) multiplied the number of application software, leading to a wave of innovation that in near future will produce robots for all commercial, industrial and consumer tasks (Ford, 2015). Yet, Hughes (2016, p. 15) highlights

“every major development in the computer’s history arose from voluntary initiative or public funding rather than corporate research”.

The nature and conceptualisation of information society is difficult to define and have been debated over the decades, being defined by Daniel Bell as a post-industrial society based on service economy rather than good-producing (Salvaggio, 2013). Webster (2006) argues most of the definitions about information society are focused on the features rather than the operational criteria, failing to understand in which ways information is becoming more relevant and why. Also, he recommends five approaches to define an information society, each one establishing proper criteria to identify the new: technological, economic, occupational, spatial and cultural. Castells (1996, p. 5) ponders, even if the *“technology does not determine society”*, it *“embodies the capacity of society to transform themselves”*, changing the destiny of economies, militar power and social well-being. Ultimately, Arthur (2009) reflects technology is expected to make people’s lives better, solving people’s problems, avoiding predicaments and providing the desired future for all and future generations.

Floridi (2016) argues the Organisation for Economic Cooperation and Development (OECD)'s criteria to measure the mature information society - based on 4 four areas: broadband and telecom, internet economy, consumer policy and digital government - should be more qualitative and includes people's expectations. OECD (2017) reports AI as a trending innovation subject that will produce fundamental transformation for business, government, and society in this new decade, with unknown opportunities and challenges regarding information security, privacy, and employment. Precisely, the report defends the AI technology will help humans to make better decisions by embedding very well tailored algorithms to previous data and settled objectives. Besides, the report recommends the creation of new indicators to track the growth and impact of the new emerging trends. Conitzer et al (2017) state the development of AI algorithms have achieved good results in the decision-making process. However, in relation to moral dilemmas it still needs a general framework to address its variety of context. Taddeo and Floridi (2018) argue AI is more than a technology in the need of regulation, it is also a strong force that is re-molding practices, interactions and environment in personal and professional levels, raising new ethical challenges.

Cath et al (2017) state the US, UK and EU governments have addressed artificial intelligence issues in three reports, published in 2016. Each report independently approached the need for developing new policies to build a 'Good AI society'. White House Office offers a positive tone, defining AI as a technology able to improve human capabilities, instead of replacing human beings. Also, the North-American report defends a limited regulating role from the government and more research to monitor AI development, focused on its accountability and transparency. The UK House of Commons report uses a more urgent tone to recommend

prevention, mitigation and governance of AI but still with a limited governmental interference. On the other hand, EU report is mostly focused on embodied AI technologies and the ethical concerns about the level of autonomy and impact of robotics in human jobs. COMEST (2019) published a preliminary study about AI ethics, addressing the most relevant ethical issues from UNESCO's perspective, which includes diversity, gender equality, education, scientific knowledge, communication, information and global sustainability.

METHODOLOGY

The philosophical approach of the research will be the interpretivism and the phenomenological paradigm to look over our established understanding of the AI ethics and to cover the current experience to discover new emerging meanings (Gray, 2014). The research will be conducted by qualitative mixed methods, with two data collection methods, comprehending focus groups and semi-structured interviews.

In order to gain a range of perspectives and listen to a more diverse group impacted by AI in society, including who could be forgotten by other approaches (Carey and Asbury, 2012), will be organised 7 focus groups, each one with 6 participants. The focus groups will take place in the UK and will be divided by ethnic minorities, LGBT community, unemployed adults, students of high education, female workers, people with disability, and a final group made by ethnic majority people between 18-65 years old. The focus groups will have the duration of 1h to discuss their daily life and concerns.

The semi-structured interviews will be conducted face-to-face in the UK and online, with 12 experts of artificial intelligence and ethics. The participants will be equally divided by stakeholder position, including government, civil society, academia, private sector and media sector. The interviews will have 10 primary questions, and will take around 30 minutes. This qualitative data collection method intends to extract the views, attitudes and meanings of the first-line people involved in the AI field (Gray, 2014).

The analysis of data will follow the thematic analysis approach, to identify, analyse and describe the most relevant patterns or themes surfaced from the research (Carey and Asbury, 2012). The research will establish the essential trust with participants to gain their consent and engagement. A not intrusive set of questions and an empathetic tone will be used in the interactions with participants to make them feel comfortable to share their thoughts. It will respect the confidential nature of the collected data, protecting the records in devices and cloud systems locked by strong passwords.

RESEARCH QUESTIONS

The literature indicates the recent development and application of Artificial Intelligence have increased in the last decade, which raised the ethical concerns in society, starting from the controversial purpose to pursue human intelligence until a big variety of applications in public and private sectors. The government of the most mature information societies, such as the US, UK and EU, and non-profit organisations like UNESCO are currently investigating or developing new frameworks and regulations to mitigate the impact of AI. It was noticed the approach of governance and accountability could change secondly the country or the economic region, being the US the less inclined to create new forms of governmental regulations. However, a global approach is required to address the main ethical issues around the world. In the meantime, the multinational corporations and SME companies are keeping their AI development and implementation with small costs, low control barriers, controversial technological relevance and high risks for the sustainability of the business in the long-term. The risk for information societies regards multiple stakeholders, industrial sectors, and ethical perspectives, with interconnected and complex consequences.

In accordance with the literature presented in this proposal, the PhD research aims to investigate the following research questions:

- What are the ethical concerns regarding artificial intelligence?
- How Artificial Intelligence ethical issues impact information societies?
- How Artificial Intelligence ethics is addressed in different information societies?

- How should the governments change the regulation of artificial intelligence to mitigate ethical issues?
- Which approach should be applied to address artificial intelligence ethical issues?

REFERENCES

Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. New York: Simon and Schuster.

Carey, M.A. and Asbury, J. (2016). *Focus Group Research*, 1st edn. New York: Routledge.

Castells, M. (1996). *The Rise of the Network Society*, 1st edn. Malden: Blackwell.

Cath, C., Wachter, S., Mittelstadt, B. , Taddeo, M. and Floridi, L. (2017). ‘Artificial Intelligence and the ‘Good Society’: the US, EU, and UK approach’, *Science and Engineering Ethics*, 24, 505–528.

Conitzer, V., Sinnott-Armstrong, W., Borg, J.S., Deng, Y. and Kramer, M. (2017). ‘Moral Decision Making Frameworks for Artificial Intelligence’, Duke University, February, 2017. Durham: Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17).

COMEST (2019). *Preliminary Study On The Ethics of Artificial Intelligence*. Available online: <https://unesdoc.unesco.org/ark:/48223/pf0000367823> [Accessed 14 July 2020]

Dreyfus, H., Dreyfus, S. E. and Athanasiou, T. (1986). *Mind Over Machine*. New York: Simon and Schuster.

Fjelland, R. (2020). 'Why general artificial intelligence will not be realized', *Humanities and Social Sciences Communications*, 7(1), 10.

Floridi, L. (2010). *The Cambridge Handbook of Information and Computer Ethics*, 1st edn. Cambridge: Cambridge University Press.

Floridi, L. (2016). 'Mature Information Societies: a Matter of Expectations', *Philosophy & Technology*, 29(1), 1–4.

Ford, M. (2015). *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books.

Gray, D.E. (2014). *Doing Research in the Real World*, 3rd edn. London: SAGE.

Heikkinen, T.H. (2019). How does the use of artificial intelligence affect the concept of fair trial?. Masters level. Lund University.

Hughes, B. (2016). *The bleeding edge: Why technology turns toxic in an unequal world*. Oxford: New Internationalist.

Kotler, P. (2015). *Confronting Capitalism: Real Solutions for a Trouble Economic System*, 1st edn. New York: AMACOM.

Lanz, P. (2000). 'The Concept of Intelligence in Psychology and Philosophy'. In H. Cruse, J. Dea and H. Ritter (eds), *Prerational Intelligence: Adaptive Behavior and Intelligent Systems Without Symbols and Logic*, pp. 19-30. Dordrecht: Springer.

Lebow, R. N., Schouten, P. and Suganami, H. (2016) *The Return of the Theorists: Dialogues with Great Thinkers in International Relations*. Basingstoke: Palgrave Macmillan.

Linden, A. and Fenn, J. (2003). 'Understanding Gartner's hype cycles' Strategic Analysis Report N° R-20-1971. Gartner, pp. 88.

OECD (2017). *Measuring the Information Society Report*. Available at <https://www.oecd-ilibrary.org/content/publication/pub-80f52533-en> [Accessed: 15 July 2020]

Penrose, R. (1994). *Shadows of the Mind*. Oxford: Oxford University Press.

Rifkin, J. (2014). *The Zero Marginal Cost Society*, 1st edn. New York: Palgrave Macmillan.

Sarmah, S.S. (2019). 'Concept of Artificial Intelligence, its Impact and Emerging Trends', *IRJET*, 6(11), pp. 2164-2168.

Salvaggio, J. L. (2013). *The information society: Economic, social, and structural issues*. New York: Routledge.

Shed, S. (2020). Researchers: Are we on the cusp of an ‘AI winter’?. BBC News [Online], 12 January. Available at <https://www.bbc.co.uk/news/amp/technology-51064369> [Accessed: 31 March 2020]

Taddeo, M. and Floridi, L. (2018). ‘How AI can be a force for good’, *Science*, 361(6404), pp.751-752.

Trend, D. (2001). *Reading Digital Culture*, 1st Edn. Oxford: Blackwell.

Webster, F. (2006). *Theories of The Information Society*, 3rd edn. Oxon: Routledge.

Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgement to Calculation*, 1st edn. San Francisco: Freeman.