Do we really care about Artificial Intelligence? A Review on Social Transformations and Ethical Challenges of AI for the 21th Century

João Guerreiro

Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL), Lisboa, Portugal
Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Lisboa, Portugal

ABSTRACT

Although Artificial Intelligence (AI) is based on research from the 20th century, only recently computation and new algorithms allowed AI to gain momentum and practical applications in society. Examples of such uses include self-driving cars and autonomous robots that are changing society and how we interact. However, despite this advances, the discussion about the social transformations and ethical implications of this new reality are still scarce. The current chapter reviews the current stance of ethical and social transformation discussions on AI and presents a framework for future developments. The main contributions of the chapter allow researchers to understand the major gaps in research that may be explored further in this topic and allow practitioners to gain a better picture of how AI may change society in the near future and how we should prepare for those changes.

Keywords: Artificial Intelligence, Artificial General Intelligence, Ethics, Human-Computer Interaction, Morality, Singularity, Regulations, Transparency

INTRODUCTION

Artificial Intelligence (AI) discussions have been around for years. The Alan Turing test dates back to the 1950's and was suggested as a way to determine if a machine is able to think and behave with an intelligence that resembles that of the human being (Turing, 1950). Since then AI has evolved into a robust science with advanced algorithms that learn new skills with the help of the surrounding environment. Despite such advances, AI has remained in the realms of academia and with a limited number of applications in the companies. However, recently, with the development machine learning and deep learning (Krizhevsky and Hinton, 2012), AI has moved from a research oriented science to an applied one, with many applications

that benefit society. For example, AI has been applied successfully to decision-support systems to predict fraud (Kültür and Çağlayan, 2017), and consumer behavior in companies (Zhong and Li, 2019), or to help predict medical diagnosis (Belić et al., 2019). Although former applications are mostly supervised by human beings, today AI has evolved to more autonomous decisions, albeit still controlled by humans. AI is starting to affect our daily routines, either through the use of intelligent algorithms that curate the Internet information (e.g. music in Spotify) and present information tailored to consumers' needs or in autonomous systems such as self-driving cars by improving driving skills to a point where the number of errors is limited. AI is also being used as a customer relationship tool. For example, AI has powered speakers such as Google Home or Amazon Echo that handle daily activities such as ordering a pizza or turning on the lights. AI is also expected to become more pervasive as new forms of human interactions with machines increase. The increasing use of Virtual Reality and Augmented Reality may soon be coupled with the help of AI agents that guide the virtual and augmented experience. Such growth shows that consumers are willing to let some of their privacy be shared with autonomous systems in exchange for convenience and performance (TechCrunch, 2019).

Indeed, AI applications are expected to increase the worldwide productivity by 5.5% up until 2030 (Statista, 2019a). The market of autonomous cars is also expected to rise up to 6 billion dollars in 2025 and in the healthcare sector, the growth is even more expressive. The global market size for AI in healthcare is forecasted to grow up from 1 billion (in 2016) to more than 28 billion dollars in 2025 (Statista, 2019b).
The predictions of growth are aligned with citizen's trust in AI. For example, around 62% of U.S. consumers are open to use AI to improve their experiences (Statista, 2019c) and in 2018, 70% of Chinese people reported finding AI trustworthy (Statista, 2019d). Therefore, developing AI systems is not only a vague promise – as in the 20th century - but a reality that is set to change the future of mankind in the next future.

The power of AI has grown at fast pace since the beginning of the 21th century. The advances and predictions for the future both in terms of learning techniques and applications of AI systems have led governments, computer scientists, philosophers and managers to start developing a set of discussions about the implications of an AI society in the future (Nebeker et al., 2019). Concerns such as lack of data privacy, biases in the learning algorithms, lack of AI accountability and implications for labor are driving the agenda. However, despite the recent regulations and discussion forums, the ethical issues and moral dilemmas that AI systems will face as they grow in intelligence are still to be fully understood. Although businesses and consumers well-being may benefit from introducing AI into their daily activities, AI agents must comply with different ethics, standards and regulations so that they protect humankind without compromising their own development.

The current chapter reviews the current stance of ethical and social transformations of AI by conducting a literature review of papers written in the 21th century about ethical principles that should be applied to AI systems. A text mining approach is used to identify the main topics covered in the literature and the most relevant papers in each topic. Solutions and recommendations for future research are then suggested in order to set new challenges for the next decades.

BACKGROUND

The rise of the machines has been a common theme in fictional novels throughout time. Movies such as 2001: A Space Odyssey, The Terminator and Ex Machina have always characterized the relationship between humans and artificial intelligence as tense and conflictuous. However, the exponential growth of AI in the 21th century has made such coexistence a reality. Although its emergence has been confined to specific applications, we all remember how Deep Blue from IBM won the world chess champion Gary

Kasparov in 1997 or more recently (2016) how AlphaGo beated Lee Sedol in one of the most complex, intuitive and creative board games in the world – Go (AlphaGo, 2019). Nowadays, Artificial Intelligence is starting to have some implications in our daily lives. Companies are using AI to build more advanced interactions with their consumers through the use of Virtual Reality (Luck and Aylett, 2000), smart speakers and chatbots (Angeline et al., 2018), to uncover fraudulent activities and stop them before they may affect the business (Kültür and Çağlayan, 2017), to create smarter cities, or even to place AI in autonomous cars or military drones (Elliot, 2019). The possibility of AI evolving to a singularity or a state of Artificial General Intelligence (AGI) is much closer than in the end of the 20th century.

The potential applications of AI are vast, and so are the implications for society depending on the level of the artificial intelligence. According to Huang and Rust, (2018), AI may be divided into four different types: (1) Mechanical, which perform tasks that require limited training or education, (2) Analytical, which include all types of machines and algorithms trained for particular tasks and that may adapt and decide on a specific field (for example using a neural network trained to predict the probability of a consumer to leave the company, (3) Intuitive, which include AI systems able to think in abstract terms, for creative learning, and to solve new challenges (e.g. autonomous vehicles) and finally (4) Empathetic, which are systems that go beyond human intellect and may adapt empathically, communicate and learn with others – sometimes also referred as singularity or AGI (e.g. humanoid robot Sophia). Having mechanical artificial agents performing repetitive tasks for us may lead to more satisfying jobs – albeit it may cut some jobs as well (Brynjolfsson, 2014). Intuitive AI may solve the most important challenges that the humankind faces today, such as, global climate change and space exploration. However, if AI becomes more empathetic and intelligent it may compete with human intelligence and even create new social classes, which may lead to imbalances in our fragile society. Therefore, many scholars, governments and private companies have been tackling the philosophical issues that may arise due to such coexistence.

The moral restrictions embedded in the human society – and that are still little understood (Shulman et al. 2009; Bello and Bringsjord 2012; Bostrom 2014; Brundage 2014) – help us overcome moral dilemmas and find heuristics that guide decisions. Human researchers, for example, know that the cure for cancer cannot be handled by eliminating every human being on earth, but for an amoral algorithm that may be a possible solution (Alotaibi, 2018). Therefore, as AI becomes more pervasive and ubiquitous in society, there is a need to set rules, standards and ethical boundaries that may be used in the future. The first rules for AI coexistence with the human kind date back to the same time Alan Turing suggested his test for artificial intelligence. In the 1950s, Asimov (1950) suggested a set of 3 rules that were later expanded to include a zero law (Asimov, 1985). The rules were focused on robots – a specific kind of machine that may be driven by artificial intelligence. However, nowadays, AI may be embedded in many devices other than robots, such as smartphones, virtual/augmented reality headsets, cars, IoT devices, pacemakers, and others. According to Asimov (1950, 1985):

0. "A robot may not injure humanity or, through inaction, allow humanity to come to harm."
1. "A robot may not injure a human being or, through inaction, allow a human being to come to harm, unless this would violate the Zeroth Law."
2. "A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law."
3. "A robot must protect its own existence as long as such protection does not conflict with the Zeroth Law, the First Law or the Second Law."

Despite the clear and simple instructions set forward by Asimov, researchers have learned that such simple instructions are not enough to guide an artificial intelligence system that may learn to define its own rules, and especially that needs to make decisions with moral dilemmas. For example, an AI system embedded in a self-driving car may have to decide to avoid a child in the middle of the road by sacrificing the passengers in the car. An AI system embedded in a human brain used to improve the human well-being and

performance may have to decide/act on issues that may break implicit ethical rules in society such as corruption or ethnical/gender bias. Decisions such as these are not easy to take, even for humans and therefore, there is a need to come up with a better way to print such constraints into AI systems as well. Therefore, many groups and forums have recently been assembled to present regulations and guides for designing ethical AI systems. Table 1 shows the most active initiatives thus far.

Table 1. Ethical Initiatives for AI Systems

Program	Goal	Reference
Asilomar Principles for Beneficial AI	Develop Best Practices	https://futureoflife.org/ai-principles/
IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	Develop Standards, Certifications and Codes	https://standards.ieee.org/industry-connections/ec/autonomous-systems.html
Partnership on AI to Benefit People and Society	Develop Best Practices	https://www.partnershiponai.org/
Barcelona Declaration for the Proper Development and Usage of AI in Europe	Develop Best Practices	Steels and Mantaras (2018)
AI-100	Impact on AI on urban life by 2030 in North America	https://ai100.stanford.edu/
AI Now Institute	Conduct evidence-based research	https://ainowinstitute.org/
Human Rights, Big Data and Technology Project	Analyze the use of big data, artificial intelligence, associated technologies	https://hrbdt.ac.uk/
High-Level Expert Group on Artificial Intelligence	Recommend ELSI policy development on AI	https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
Chinese Association for Artificial Intelligence	Unite artificial intelligence science and technology professionals	http://www.caai.cn/
AI for Humanity	Create an international group of AI experts to prepare for societal transformation	https://www.aiforhumanity.fr/en/

Souce: (*adapted from Nebeker et al., 2019 and Steels and Mantaras, 2018)*

Such recommendations are mainly based on a two way approach. A first bottom-up approach that is based on developing AI systems that may be able to learn how to distinguish between moral and immoral behaviors from past actions (Bello and Bringsjord, 2012) and a top-down approach based on moral rules that are embedded in AI systems (Arkoudas et al., 2005; Oesterheld, 2016). A third hybrid approach has also been proposed by Wiltshire (2015) that uses the advantages of both top-down and bottom-up approach (Bogosian, 2017). More recently, Floridi et al. (2018) suggested that AI systems may have to be constrained under a *soft ethics* and *hard ethics* standards. *Hard ethics* are those values, rights, duties and responsibilities that are implicit in society and that must regulate AI in a top-down approach. *Soft ethics* are those that have been implemented by regulations, such as those imposed by the EU on data privacy (GDPR) – a bottom-up approach. Usually, *hard ethics* contribute to generate the *soft ethics* regulations and standards.

MAIN FOCUS OF THE CHAPTER

The main focus of the current chapter is to summarize the literature review on ethics in AI and to present a set of future research directions. In order to uncover the latest ethical issues that have been addressed in the

literature, a collection of relevant literature was extracted from Web of Science using the following query applied to the title, abstract and keywords: "artificial intelligence" and ethic*. The use of wildcards on ethical issues was used so that results returned terms such as ethics, ethical and other related terms. The query was restricted to only papers published on peer-review journals published in English language.

A total of 322 papers were retrieved from WoS, from which 145 were selected for further analysis due to having ethic related keywords. A final systematic analysis based on Nill & Schibrowsky (2007), Moher, Liberati, Tetzlaff, Altman & Altman (2009) and Loureiro et al., (2018) was used to filter out papers that were not addressing ethical issues as their core subject. Two researchers discussed the final selected papers and reached an agreement > 0.85 in terms of Cohens' Kappa coefficient. A final set of 71 papers were agreed to be relevant for the current study and their abstract was extracted for further analysis.
After converting text into lowercase and removing any special characters from text, sentences were split into tokens using a regular expression ("\w+"). A normalization procedure was used to guarantee that words that have the same meaning were coupled together. A *UDPipe* Lemmatizer was used to achieve such normalization (Straka and Straková, 2017). The sentences were also stripped from any *stopwords* in English, including regular words that although are common in the text do not add much information to the final analysis such as "artificial intelligence" (Guerreiro et al., 2016).

Results reveal that most papers were published recently (2018 and 2019) while only one paper addressed specifically ethics in AI in the year 2000. Figure 1 shows the distribution of papers over the years.

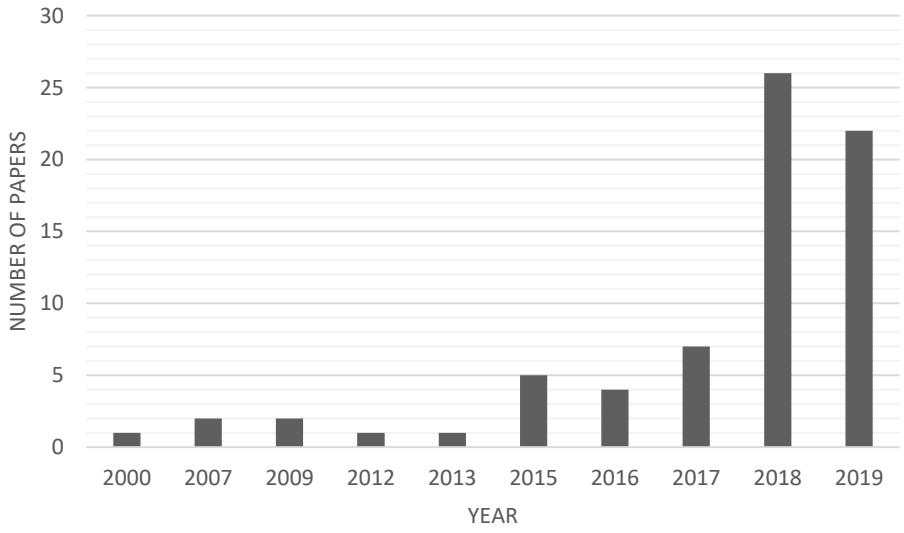

Figure 1. Number of papers on Ethical discussions over the 21th Century

A text mining approach based on a Latent Dirichelet Allocation (LDA) algorithm was used to uncover latent topics in text (Blei et al., 2003). Table 1 shows the topics extracted by using LDA algorithm along with the terms more correlated with each topic and papers highly correlated with each group ordered by posterior probability of belonging to each topic.

Table 1. Latent Topics and correlated papers

Topic Name	**Topic Terms**	**Correlated Papers with the Topic**	**Post. Prob.**
T1. DESIGN OF AI SYSTEMS: Risks and Challenges	approach, design, principle	McKernan et al. (2018) Arnold et al. (2019) Sotala and Yampolskiy (2015)	.802 .736 .720

T2. MORALITY IN AI	moral, agent, development	Roff (2019) Bagosian (2017) Bryson (2018)	.763 .566 .564
T3. HUMAN-MACHINE INTERACTION	human, machine, robot	van der Meulen and Bruinsma (2019) Nebeker, Tourous and Elis (2019) Livingston et al. (2019)	.953 .868 .852

TOPIC ANALYSIS

T1. DESIGN OF AI SYSTEMS: RISKS AND CHALLENGES

Approaches towards the design of AI systems have been discussed in the literature in recent years. Such approaches are focused on AI as both an opportunity and a cost (liability) that must be handled with care. Although AI may enable self-realization, the literature has pointed out a need to create a "smart agency" that may regulate AI in terms of transparency, responsibility and accountability, without which, human self-determination may be at risk (Floridi et al., 2018). The work of Clarke (2019) suggests a set of principles to guide organization businesses in developing AI applications. The author proposed 10 themes that must be regulated to develop responsible AI technologies, artefacts, systems and applications, namely: (1) Assess positive and negative impacts and implications, (2) complement humans, (3) ensure human control, (4) ensure human safety and well-being, (5) ensure consistency with human values and human rights, (6) deliver transparency and auditability, (7) embed quality assurance, (8) exhibit robustness and resilience, (9) exhibit accountability for obligations and (10) enforce and accept enforcement of liabilities and sanctions.

AI is often addressed in the literature in terms of its risks, such as the (1) risk of being unsafe – for example prone to be hacked, (2) lacking transparency, (3) potentially biased and unfair, (4) inducing unemployment and moral de-skilling, (5) creating a socio-economic inequality, (6) dependency and (7) having a potential effect on human relationships (Green, 2018). However, McKernan et al. (2018), also discusses the design of AI as a compassionate and transparent system that may help society – in the specific case to help prevent suicides. The authors suggests a framework supported on three recommendations – consent, control and communication. Indeed, consent is fundamental to ensure that AI systems are used properly. Due to the pervasiveness of AI, the risks of having such systems affecting citizens and consumers lives without their consent is a risk that must be addressed. Control refers to AI oversight and transparency and finally communication allows the system to be used not only by a small part of the population but more broadly.

Sotala and Yampolskiy (2015) discusses the emergence of AGI (Artificial General Intelligence) and how it may pose risks and challenges to society, mainly by its ability to go beyond human intelligence. The authors make a review around the main literature on AGI and discuss some proposals on how society should deal with the emergence of AGI, namely by (1) letting the AGI system regulate itself, (2) integrate AGI into society and develop legal and economic controls, (3) regulate research on AGI to encourage international compliance and safety research or (4) restrict the ability for humans to develop super intelligent agents.

A final paper also highly correlated with this topic addresses the implementation of a "big red button" into the design of AI systems (Arnold and Scheultz, 2018). The main underlying idea of a red button to shut down AI systems that do not conform to the ethical principles of humankind is supported under the assumption that although explicit ethical rules may exist, many AI systems will not be aware of such normative principles and, therefore, are prone to fail (Moor, 2006). In fact, as AI becomes more pervasive and general, intelligent agents should communicate between them. If not all of them share the same rules and regulations there may be a need to handle such inconsistency. Hence, Arnold (2019) suggests a set of ethical tests that should be enforced into the AI systems so that they may terminate their activities even before a potential hazardous effect may occur. The authors propose a safety valve that may be used in

scenarios of moral dilemmas or when the system may act contrary to the society best interests in order to improve performance.

T2. MORALITY IN AI

Despite the agreement that AI systems must comply with a set of moral norms that regulate how they operate, the main problem so far has been the lack of understanding of what those moral norms should be (Shulman et al. 2009; Bello and Bringsjord 2012; Bostrom 2014; Brundage 2014). Such moral disagreement has been one of the main challenges in trying to reach a consensus on how to deal with machine Ethics. However, researchers have tried to come up with a set of models of a single theory that may be applied to AI systems (Bogosian, 2017). Top-down models have been proposed based on the works of Immanuel Kant, for which moral actions are based on rationality and personal freedom and from David Hume, who argued that reason was not a way to guide moral values (Livingston, 2019). However, many argue that a Kantian view of AI is not a proper way of establishing a moral agent given that AI is bounded in its ability to be free (Tonkens, 2009). Other theories for designing moral agents have stem from utilitarianism (Oesterheld, 2016) which suggest a utility function that could be used to check moral dilemmas.

Although Roff (2019) argues that it is not possible to develop a moral AI system given that it is constrained by its ability to learn from past data, being unable to make assumptions from unseen data, Bryson (2018) suggests that AI may be developed to recognize socially-acceptable actions and conform to such norms from the past in order to inform their decisions in the future (Bryson, 2018; Cakmak et al. 2010; Riedl and Harrison 2015). The question that remains is how to code what is deemed socially acceptable or not, and how can such a system identify a bias in society that is morally condemned. Further, Bryson (2018) explores another important question of AI: accountability. Who is liable for the actions of an AI agent that is working for a specific company? The AI agent that breaks the moral norms or the human that is behind its learning?

Another paper highly correlated with this topic addresses the complexity of incorporating emotions in the robot perceptions to help guiding such moral behaviors (Stowers et al., 2016). The authors suggest four main guidelines for designers, namely: (1) to codify human morality from past moral behavior (2) to codify robot morality – embedding a moral code into robot decisions, (3) implement awareness in robots, including the notion of self-awareness and finally, (4) govern the use of robots – defining a set of international regulations and laws that govern artificial intelligence systems.

T3. HUMAN-MACHINE INTERACTION

The discussions around Human-Machine interaction (HMI) have been vast in the last decades. The discussion around HMI may go from the simple interaction of a machine with a person in the form of a smart speaker, to Brain-Computer interfaces (HCI) that embed AI systems into the Human brain through neuro-stimulators and neuro-prosthetics (Glannon, 2016). The possible augmentation of the human body through the use of technology conforms with the theory of the Extended Mind of Clark and Chalmers (1998) and with the ideals of Transhumanism set forward by Bostrom (2005a, 2005b) in which humans are destined to integrate technology to evolve to higher levels of performance and intelligence. More recently, van der Meulen and Bruinsma (2019) discussed the move from the individual (human) to a datavidual (man as aggregate of data). The authors reflect on the ethical principles that must guide the evolution from the *homo sapiens* to a *homo technologicus* (Zehr, 2011; Zehr, 2015), based on a deeper awareness, knowledge and sense of critique of AI so that humans are not trapped in the "*straitjackets*" of AI algorithms. In sum, the authors proposes that AI systems that are integrated in humans should not be allowed to decide and make considerations about good and evil so that free will and human morality is not dissolved into technology. Therefore, such argument restricts the possible types of AI systems that may be allowed to be developed for Human-Machine interactions, namely when it comes to Brain-Computer interfaces. Such challenges

require research to go further on detailing when and how that may occur, and international rules to define the ethical limits of such use.

In the current topic, Nebeker, Tourous and Elis (2019) also discusses how human-machine interaction must be regulated in the healthcare sector. The authors suggest a framework based on ethical principles such as respect for the persons involved, beneficence and justice. Such principles must be used to regulate data collection and use for the purpose of developing AI systems that learn from experience, privacy, accessibility and risks that may emerge from using the technology.

Finally, the paper of Livingston and Risse (2019) is also an important discussion to understand the future impact of AI on Humans and Human Rights. The authors discuss how AGI will change the future of mankind for example in empowering humans to go beyond their cognitive and physical limitations, but also by producing a shift in political authority, namely by the emergence of new social classes that may produce an imbalance in the very fragile society we live today.

FUTURE RESEARCH DIRECTIONS

The current chapter reviewed the main concerns that are being addressed today by governments, researchers and companies to help overcome the potential risks of a non-regulated AI society. However, much is yet to be discussed from a more philosophical to a more applied perspective. One of the biggest challenges is still to code the Human moral norms into a set of rules that may be embedded into AI systems. Although such task must be accomplished by multi-disciplinary stakeholders, it is a fundamental basis for allowing AI to become pervasive and autonomous in its decisions. Continuing from the current state of the art depicted in the last section we here present on Table 2 a set of research questions in AI ethics that we suggest may be addressed by researchers and practitioners in the next decade.

Table 2. Future Research Question about Ethics on AI

Topic	Future Research Questions
Design Of Ai Systems: Risks And Challenges	How to design AI systems that are able to do creative jobs without replacing human creativity? How to design AI systems that incorporate cultural diversity and company values into their decisions? How to design unbiased AI systems even when learning from biased data (e.g. racial, gender bias)? How to design AI systems that are able to enhance humankind without subverting its principles?
Morality In AI	What moral codes should guide AI systems? Should this moral codes be internationally established or regional (adapted to each culture and values)? How far can AI systems be accountable for their actions?

	How to control and evaluate moral behaviors of AI systems? What issues may emerge when Humans allow AI to have freewill to decide and at the same time use it to enhance the Human society at the cost of a servant AI society? If there is a red-button to prevent an AI system to fail, how to prevent the same intelligent system to ignore such instructions in order to survive?
Human-Machine Interaction	How can AI systems be used to enhance human/consumers performance, engagement and well-being? Should there be any ethical limits to human-machine integration? How to avoid AI driven technology embedded in a human brain to avoid controlling human freewill? How to prevent *homo sapiens* to be replaced by *homo technologicus*?

CONCLUSION

For the last 70 years Artificial Intelligence have been evolving through a succession of advances and setbacks. Ultimately, AI has evolved to become a robust field of knowledge able to make autonomous decisions. The future of AI is still being shaped, but in the next decades we may witness, for the first time in the history of mankind, the emerging of a super-intelligent system that surpasses human intelligence. Although we are still far from reaching such an Artificial General Intelligence, the design of smart systems that today rely on AI to drive autonomous cars, to manage smart cities, to establish relationships with consumers, demand that we start defining how to deal with ethical issues in the future. As intelligent systems become pervasive in our society, they will also test our ability to teach a new "*species*" how to behave with a set of moral norms and regulations. However, there is still a long road ahead in establishing the frameworks that model our society.

The current chapter presented a literature review on the core papers addressing ethics in artificial intelligence during the 21th century and presents three major topics that have been discussed so far. We then proposed a set of questions that are still being debated and need an urgent answer from researchers, managers and governments. Such questions may be used as a starting point for further debates and studies to explore how AI may affect our civilization – economic model, sustainability, labor, healthcare, well-being, and political models. The social impacts of not regulating the evolution of AI will have big implications for citizens, consumers, companies and states.

REFERENCES

AlphaGo (2019). *AlphaGo Korea – DeepMind.* Retrieved from https://deepmind.com/alphago-korea

Alotaibi, S. S. (2018). Ethical issues and related considerations involved with artificial intelligence and autonomous systems. *International Journal of Advanced Computer Science and Applications, 9*(4), 35–40.

Angeline, R., Gaurav, T., Rampuriya, P., & Dey, S. (2018). Supermarket Automation with Chatbot and Face Recognition Using IoT and AI. In *2018 3rd International Conference on Communication and Electronics Systems (ICCES)* (pp. 1183-1186).

Arnold, T., & Scheutz, M. (2018). The "big red button" is too late: an alternative model for the ethical evaluation of AI systems. *Ethics and Information Technology, 20*(1), 59–69.

Asimov I (1950) "*Runaround". I, Robot.* New York, NY: Street & Smith.

Asimov, I. (1985). *Robots and empire*. New York: Doubleday.

Belić, M., Bobić, V., Badža, M., Šolaja, N., Đurić-Jovičić, M., & Kostić, V. S. (2019). Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease–A review. *Clinical neurology and neurosurgery*, 105442.

Bello, P., & Bringsjord, S. (2012). On how to build a moral machine. *Topoi, 32*(2), 251–266.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.

Bostrom, N. (2005a). A history of transhumanist thought. *Journal of Evolution and Technology, 14*(1), 1–25.

Bostrom, N. (2005b). Transhumanist Values. *Journal of Philosophical Research, 30*(supplement), 3–14.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.

Brundage, M. (2014). Limitations and risks of machine ethics. *Journal of Experimental & Theoretical Artificial Intelligence, 26*(3), 355–372.

Brynjolfsson, E. (2014). The second machine age : work, progress, and prosperity in a time of brilliant technologies. In A. McAfee (Ed.), *Work, progress, and prosperity in a time of brilliant technologies* (1st ed.). New York: New York W. W. Norton & Company.

Bryson, J. J. (2018). Patiency is not a virtue: the design of intelligent systems and systems of ethics. *Ethics and Information Technology, 20*(1), 15–26.

Cakmak, M., Chao, C., & Thomaz, A. L. (2010). Designing interactions for robot active learners. *IEEE Transactions on Autonomous Mental Development, 2*(2), 108–118

Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law and Security Review, 35*(4), 410–422.

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis, 58*(1), 7-19.

Elliott, A. (2019). Automated mobilities: From weaponized drones to killer bots. *Journal of Sociology*, *55*(1), 20-36.

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, *6*(1), 205395171986054.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines, 28*(4), 689–707.

Glannon, W. (2016). Ethical issues in Neuroprosthetics. *Journal of Neural Engineering*, 13(2), 1-22.

Green, B. P. (2018). Ethical reflections on artificial intelligence. *Scientia et Fides, 6(*2), 9–31.

Guerreiro, J., Rita, P., & Trigueiros, D. (2016). A text mining-based review of cause-related marketing literature. *Journal of Business Ethics*, *139*(1), 111-128.

Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage.

Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155-172.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Kültür, Y., & Çağlayan, M. U. (2017). A novel cardholder behavior model for detecting credit card fraud. *Intelligent Automation & Soft Computing*, 1-11.

Livingston, S., & Risse, M. (2019). The Future Impact of Artificial Intelligence on Humans and Human Rights. *Ethics & International Affairs*, *33*(2), 141-158.

Loureiro, S. M. C., Guerreiro, J., Eloy, S., Langaro, D., & Panchapakesan, P. (2019). Understanding the use of Virtual Reality in Marketing: A text mining-based review. *Journal of Business Research, 100*, 514-530.

Luck, M., & Aylett, R. (2000). Applying artificial intelligence to virtual reality: Intelligent virtual environments. *Applied Artificial Intelligence, 14*(1), 3-32.

McKernan, L. C., Clayton, E. W., & Walsh, C. G. (2018). Protecting Life While Preserving Liberty: Ethical Recommendations for Suicide Prevention with Artificial Intelligence. *Frontiers in Psychiatry, 9*(December), 1–5.

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems, 21*(4), 18–21.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Altman, D. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 8*(7716), 336–341.

Nebeker, C., Torous, J., & Bartlett Ellis, R. J. (2019). Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Medicine, 17*(1), 137.

Nill, A., & Schibrowsky, J. A. (2007). Research on marketing ethics: A systematic review of the literature. *Journal of Macromarketing, 27*(3), 256–273.

Oesterheld, C. (2016). *Backup utility functions as a fail-safe AI technique.* Retrieved from https://foundational-research.org/files/backup-utility-functions.pd

Riedl, M. & Harrison, B. (2015). Using stories to teach human values to artificial agents. In *AI, Ethics, and Society: Workshop at the Thirtieth AAAI Conference on Artificial Intelligence*.

Roff, H. M. (2019). Artificial Intelligence: Power to the People. *Ethics & International Affairs, 33*(2), 127–140.

Saygin, A. P., Cicekli, I., & Akman, V. (2000). Turing test: 50 years later. *Minds and machines, 10*(4), 463-518.

Shulman, C., Jonsson, H. and Tarleton, N. (2009), "Machine ethics and super intelligence", In *The Fifth Asia-Pacific Computing and Philosophy Conference, 1-2 October, University of Tokyo, Japan*, (pp. 95-97).

Sotala, K., & Yampolskiy, R. V. (2015). Corrigendum: Responses to catastrophic AGI risk: A survey. *Physica Scripta, 90*(6), 1-33.

Statista (2019a). *GDP Gain from AI by Industry Sector 2030.* Retrieved from https://www.statista.com/statistics/941769/gdp-gains-from-ai-by-industry-sector/

Statista (2019b). *Healthcare AI Market Size 2015 global forecast.* Retrieved from https://www.statista.com/statistics/826993/health-ai-market-value-worldwide/

Statista (2019c). *Customer Attitudes towards AI 2019.* Retrieved from https://www.statista.com/statistics/1042755/worldwide-customers-attitude-towards-ai/

Statista (2019d). *Trust in Artificial Intelligence by Country 2018.* Retrieved from https://www.statista.com/statistics/948531/trust-artificial-intelligence-country/

Stowers, K., Leyva, K., Hancock, G. M., & Hancock, P. A. (2016). Life or Death by Robot? *Ergonomics in Design, 24*(3), 17–22.

Straka, M., & Straková, J. (2017, August). Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies* (pp. 88-99).

TechCrunch (2019). *Over a quarter of US adults now own a smart speaker, typically an Amazon Echo.* Retrieved from https://techcrunch.com/2019/03/08/over-a-quarter-of-u-s-adults-now-own-a-smart-speaker-typically-an-amazon-echo/

Turing, A. (1950), 'Computing Machinery and Intelligence', *Mind, 59*(236), pp. 433–460.

van der Meulen, S., & Bruinsma, M. (2019). Man as 'aggregate of data': What computers shouldn't do. *AI and Society, 34*(2), 343–354.

Zehr, E. P. (2011). *Inventing iron man : the possibility of a human machine.* Baltimore : Johns Hopkins University Press,

Zehr, E. P. (2015). Future think: Cautiously optimistic about brain augmentation using tissue engineering and machine interface. *Frontiers in Systems Neuroscience, 9*, 1–5.

Zhong, J., & Li, W. (2019, March). Predicting Customer Churn in the Telecommunication Industry by Analyzing Phone Call Transcripts with Convolutional Neural Networks. In *Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence* (pp. 55-59). ACM.

ADDITIONAL READING

Allen, C., Varner, G., & Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental and Theoretical Artificial Intelligence, 12*(3), 251–261.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. *Science and Engineering Ethics, 24*(2), 505–528.

Howard, A., & Borenstein, J. (2018). The Ugly Truth about Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity. *Science and Engineering Ethics, 24*(5), 1521–1536.

Khakurel, J., Penzenstadler, B., Porras, J., Knutas, A., & Zhang, W. (2018). The Rise of Artificial Intelligence under the Lens of Sustainability. *Technologies, 6*(4), 100.

Köse, U. (2018). Are We Safe Enough in the Future of Artificial Intelligence? A Discussion on Machine Ethics and Artificial Intelligence Safety7. BRAIN. *Broad Research in Artificial Intelligence and Neuroscience, 9*(2), 184–197.

Martin, D. (2017). Who Should Decide How Machines Make Morally Laden Decisions? *Science and Engineering Ethics, 23*(4), 951–967.

Petersen, S. (2007). The ethics of robot servitude. *Journal of Experimental and Theoretical Artificial Intelligence, 19*(1), 43–54.

Shank, D. B., DeSanti, A., & Maninger, T. (2019). When are artificial intelligence versus human agents faulted for wrongdoing? Moral attributions after individual and joint decisions. *Information Communication and Society, 22*(5), 648–663.

Turing, A. (1969), 'Intelligent Machinery', In *D.M.B. Meltzer ed. Machine Intelligence 5, Edinburgh University Press*, pp. 3–23.

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. *Journal of Service Management, 29*(5), 907–931.

KEY TERMS AND DEFINITIONS

Artificial Intelligence, is the science of developing artificially intelligent agents that are able to pass the Turing test.

Artificial General Intelligence, refers to artificial agents that surpass the intellect of any human.

Brain-Computer-Interface, interfaces that allow people to interact with a computer (e.g. smartphone, AI system), by using the brain waves.

Internet-of-Things (IoT), is a network of connected physical world objects that are addressable and interact with an external virtual environment.

Human-Machine Interaction, are all the interactions between a human being and a computer (either a simple computer or an intelligent agent).

Smartcities, a concept of interconnected virtual and physical objects (IoT) that together are used to control, manage and improve the city sustainable development.

Transhumanism, is a movement that believes that technology may be used in a positive way to enhance the human condition.