AI and Ethics (2021) 1:173–181
https://doi.org/10.1007/s43681-020-00022-3

ORIGINAL RESEARCH

Management perspective of ethics in artificial intelligence

Josef Baker-Brunnbauer[1]

Received: 11 September 2020 / Accepted: 26 October 2020 / Published online: 16 November 2020

Abstract
This research addressed the management awareness about the ethical and moral aspects of artificial intelligence (AI). It is a general trend to speak about AI, and many start-ups and established companies are communicating about the development and implementation of AI solutions. Therefore, it is important to consider different perspectives besides the technology and data as the key elements for AI systems. The way in which societies are interacting and organising themselves will change. Such transformations require diverse perspectives from the society and particularly from AI system developers for shaping the humanity of the future. This research aimed to overcome this barrier with the answers for the question: What kind of awareness does the management of AI companies have about the social impact of its AI product or service? The central research question was divided into five sub-questions that were answered by a fundamental literature review and an empirical research study. This covered the management understanding of the terms moral, ethics, and artificial intelligence; the internal company prioritization of moral and ethics; and the involved stakeholders in the AI product or service development. It analysed the known and used ethical AI guidelines and principles. In the end, the social responsibility of the management regarding AI systems was analysed and compared.

Keywords Artificial intelligence · Ethics · Management · Social impact · Humanity

1 Introduction

This research aimed to generate the awareness of ethical challenges for artificial intelligence (AI) systems and to analyse the management perspective and understanding of ethics for its AI product or service. Ethics, based on the used framework, are not just only about defining what is right and what is wrong. As digitalisation covers many different technologies and aspects, AI can be seen as one of them that will change not only businesses, but also humanity. In addition to new technologies and use cases, AI has a deep impact on society and social life and has the potential to seriously shape and change humanity. The increasing digitalisation at all levels [12] does not only lead to the improvement and optimisation of the products and processes but also changes the way of internal and external collaboration. Companies need to increase flexibility and openness to innovate new business models with the intelligent usage of new technologies, such as AI [2]. A high level of automation with AI systems generates an improved rating of the company performance and can potentially eliminate the existing jobs and increases the psychological pressure on the employees. Digitalisation is a challenge to employees who execute tasks that are easy to automate and to middle and high management. Digital technologies have important economic and social aspects. Companies strive for product innovations and inventions with new technologies. Moreover, the long-term impacts of digital transformation and new technologies, such as AI, are not clear from the beginning. Digital products and services are often developed by computer scientists for technical use focused on revenue and growth. Social components, considering the big picture of mankind and taking social responsibility into account, do not always have a high priority for the management. Considering the considerable impact of AI systems on society, it is of high importance that companies actively prioritise their social responsibility and take actions.

In this research, the ethical aspects of AI systems from the management perspective and their social impact implications were analysed to further investigate these challenges

✉ Josef Baker-Brunnbauer
josef.baker-brunnbauer@edu.uni-graz.at

[1] Karl-Franzens University of Graz, Universitätsplatz 3, 8010 Graz, Austria

by answering the research question based on the following sub-questions:

- What types of definitions for AI, ethics, and morals are used by the management?
- What types of prioritisation does ethics and morals have for the management?
- Who are the responsible internal and external stakeholders for designing an AI product or service?
- Are there AI guidelines and principles known and used within the company?
- What type of social responsibility does the management mention in a self-reflection?

2 Ethics and morals

An AI system can be abused by somebody with a lack of morals. Western ethics are based on several attitude and responsibility frameworks, including teleology (e.g. utilitarianism, antiquity, and hedonism) and deontology (e.g. virtue ethics) [17]. Future AI systems will operate in a more integrated manner with humans and may have their own moral status, such as being their own moral entity or doing tasks by their own will [3]. Many ethical principles have social emotions, such as compassion and empathy in common. The parameters are the reward and the punishment for the guidance. If a human does something bad and also feels bad about it, an emotional punishment is generated by the brain. If a human disregards ethical principles, the society may punish through shaming by peers or sentence at the court. There is no common ethical consensus in today's world, but there are basic principles with a broad agreement [19]. In the past, human societies had ethical principles with the focus on survival. In 2006, the concept of machine ethics that was proposed by Anderson and Anderson started discussions about ethical issues. Ethics are a complicated and complex concept with a focus on a single aspect [18].

The currently used AI technology refers to narrow-AI or weak-AI, and the ethical issues involve human interaction that will even expand when future AI systems will have the ability for determining their own moral status. Therefore, an AI system should not be treated as a machine, rather as an object having similar or equal rights as humans [18]. To avoid the misuse of the AI technology, the value of an ethical approach of AI technologies needs a strong focus and compliance with an adopted law. Floridi et al. [8] defined the transformation of an ethical approach to AI ethics as 'dual advantage'. This means that one point is that ethics support organisations to take advantage of the AI-enabled social values (new opportunities which are socially acceptable). The other point is that ethics makes it possible for organisations to prevent costly mistakes. The benefit of this prevention is the possibility to eliminate actions which are socially unacceptable but legally approved. The dual advantage of ethics can only work in a setting of public trust and clear responsibilities [8]. People will accept AI systems only if their outcome is seen as meaningful with a low risk level. The success will depend on public engagement with AI technologies, openness about operation, and the ease of understanding the AI systems by humans. Defining applied ethics for cases, such as fast recursive reproduction needs completely different ethical principles for foundational normative truths, and it is important not to take the current familiar human principles as the standards.

2.1 AI ethics implementation

One of the most important factors to consider for AI algorithm training is the human bias, such as the gender bias or race bias. As AI systems need plenty of data to train with accuracy, the datasets are chosen by humans first-hand. In this process, the existing biases may be transferred to AI systems when they develop themselves for the future. Therefore, it is important to train algorithms without human biases [18]. The deep learning AI model GPT-3 from OpenAI makes decisions based on 175 billion parameters. Researchers have conducted an analysis of biases to better understand their model regarding fairness and bias. Their study showed that internet-trained models have an internet-scale bias [4]. The model tends to reflect stereotypes based on the training data of 175 billion parameters. For example, academic and higher-paid jobs were associated with male persons; Christianity was associated with ignorant, judgmental and execution; and the Islam was linked to terrorism, fasting, and Allah [4]. If AI systems get their own sentience in the future, will they generate their own biases? There are three potential ways to educate AI systems of ethics [14]:

- Implicit ethical agents: forcing the machines' actions to prevent unethical outcome.
- Explicit ethical agents: explicitly quote the allowed and the forbidden actions.
- Full ethical agents: machines have consciousness, free will, and intention.

Besides these three approaches, how to ethically interact with an AI system that has consciousness, emotion, moral sense, and feelings is still an open question [18]. Is it ethical to shut down (kill) an AI system, if it replaces human jobs? Is it ethical to use military AI systems? This is also connected to human moral values and ethics. An AI system can be a moral agent by being autonomous (machines without the direct control of another agent), being intentional (acting in a morally harmful or beneficial manner and the actions being calculated by intention), and being responsible

(machines fulfil a social role which includes the assumed responsibilities) [18]. To make the machine behaviour (moral decision-making) acceptable to humans, we need explanations of a transparent psychological decision-making process [3, 10]. One psychological challenge for AI ethics implementation is that humans apply different moral values for themselves and to others. Anderson and Anderson [1] reported that the perfect solution will be to develop AI systems that follow ideal ethical principles. This sounds easy in theory, but it is difficult to implement in a global environment. If developers would program AI systems to be harmless to humans, the systems will need to understand the meaning of harm first. This will need a global level of ethics, which includes a reduced set of information between the ethical standard makers and the AI developers [18]. A requirement for moral AI systems is that humans accept their decisions and find them reasonable. This requires integrating broad human common sense, ethics, and values for the development of AI and particularly artificial general intelligence (AGI) systems. AI algorithms which should replace human social functions must consider the following criteria during their development: transparency, incorruptibility, responsibility, auditability, and usability [3]. A future AI/AGI system, which inspects power plant software for software bugs, will need an ethical understanding of a human engineer instead of an engineered system. AI ethics differ fundamentally from the ethical disciplines of non-cognitive technologies by the following criteria [3]:

- The specific behaviour of the AI/AGI system will not be predictable.
- The safety verification of the AI/AGI system will be more difficult, as it requires one to verify what the system is trying to do (instead of safety testing from the defined behaviour in the defined operating contexts).
- Engineering must include ethics to safeguard the development of a good AI/AGI system.

AI systems gain knowledge and skills from their interaction with their surroundings [15]. The Japanese government has created special test zones 'Tokku' for robot and AI testing. The primary aim of these zones is to install safe test environments for the scientists and for the society, where these can interact together. This fits well to the virtue–ethics–morals theory [15]. Testing scenarios in open environments show new ways to tackle legal challenges, discover new risks and threats, and prevent the possible losses of control of AI systems.

2.2 Guidelines and principles

AI systems offer considerable optimisation potential in various fields, such as transportation and logistics, or even in preventing diseases and to radically reinvent the society. Therefore, humans need to understand not to generate a dependency on AI systems and to keep the ability to make the final decision by themselves. The risk is that humanity breaks up by AI systems, as it may lead to unplanned changes by the intention to make automate routines and make people's life easier [8]. Organisations have defined the values and principles that should be used for the development and deployment of AI systems and technology within societies. Principles should serve as an ethical foundation for discussions, guidelines, standards, regulations, and laws. Floridi et al. [8] focused on the commonalities and the important differences of principles based on manifestos that generate a summary of 47 principles. Overall, there is an overlap and coherence of the different approaches, which is similar to the four core principles used in bioethics [8]: beneficence, non-maleficence, autonomy, and justice. The four bioethical principles adapt to the new ethical AI challenges. The High-Level Expert Group on Artificial Intelligence (AI HLEG) prepared the Ethics Guidelines for Trustworthy AI as part of the AI strategy of the European Commission; it is based on three components [6]: lawful, ethical, and robust. Based on fundamental rights, four ethical principles and their values must be respected for the development of AI systems: respect for human autonomy, prevention of harm, fairness, and explicability. Moreover, more vulnerable groups, such as persons with disabilities, children, people with the risk of exclusion, or situations with asymmetrical power, should get involved. Developers should adopt adequate measures to defuse risks that are difficult to predict, identify, or measure. Based on ethical principles and fundamental rights, trustworthy AI can be realised by meeting seven key requirements and considering technical and non-technical methods [6]:

- Human agency and oversight: fundamental rights, human agency, and human oversight.
- Technical robustness and safety: resilience to attack and security, fall-back plan and general safety, accuracy, reliability, and reproducibility.
- Privacy and data governance: respect for privacy, quality and integrity of data, and access to data.
- Transparency: traceability, explainability, and communication.
- Diversity, non-discrimination, and fairness: the avoidance of unfair bias, accessibility and universal design, and stakeholder participation.
- Societal and environmental well-being: sustainability and environmental friendliness, social impact, society, and democracy.
- Accountability: auditability, minimisation and reporting of negative impact, trade-offs, and redress.

The European non-discrimination law differentiates between direct (illegal and less favourable behaviour) and indirect (comparison between people, disadvantages, and seemingly neutral provisions) discrimination. The scope of the European non-discrimination law includes ethnicity, gender, religion or beliefs, disability, age, and/or sexual orientation. Automated discrimination done by AI algorithms is more abstract and unintuitive, tangible, and difficult to discover than conventional forms, as there might be a lack of access to evidence [21]. Therefore, a proposed solution is that technical and legal communities are working together to enable a consistent assessment without the interpretation of the cases of automated discrimination. The European Commission [6] offered the Assessment List for Trustworthy AI (ALTAI) that can be adapted for each specific AI use case with the aim to achieve a general framework. The European Commission [7] published a white paper on AI, aiming to become a global leader in innovation in the data economy and its applications. The document aimed to optimise research, to foster the collaboration between the European member states, and to increase the investment in AI development. Moreover, it drafted a future European regulatory framework to mobilise resources to achieve 'ecosystem of excellence' along the entire value chain. The key elements of the framework create an 'ecosystem of trust' [7]. The European Union Agency for Fundamental Rights (FRA) has collected more than 290 AI policy initiatives in the EU Member States between 2016 and 2020 [9].

In June 2020, the German automotive company Continental announced its intention to develop a code of ethics for its internal development and usage of AI that is based on the EC Trustworthy AI guideline [6]. Continental argued that smart algorithms play an important role and that it sees itself as a technology company, responsible to ensure internal ethical standards for development and processes. Furthermore, AI decision-making must always be non-discriminatory, transparent, and understandable [5]. Therefore, the automotive industry can play a lead role for AI guideline implementation within Europe. The German society TÜV (technical inspection and product certification services) urged for legal guidelines for the use of AI systems in critical safety environments [20]. According to its survey, German consumers require more transparency and safety for AI applications. This consumer demand needs to get attention by the company's management for its AI system development. In all, 85% of the participants stated that AI systems should be available only after testing and certification from an independent third party. Only 17% of the interviewees said that they would trust the AI manufacturer regarding safety. Floridi et al. [8] generated a dynamic list of 20 action points in the categories assess, develop, incentivise, and support as a recommendation to policy makers for a good AI society. AI systems should be designed to decrease inequality, respect human autonomy, and increase benefits which are usable for all humans. A highlighted point is that AI systems are explicable to build trust and understand the technology. Another defined factor is the requirement of a multi-stakeholder approach to ensure that AI systems serve the society's needs. Therefore, developers, users, and rule makers need to work together in an integrated manner.

3 Research methodology

The main focus of this research was to answer the defined central research question 'What kind of awareness does the management have about the social impact of its artificial intelligence (AI) product or service?' by conducting expert interviews. The focus was on the geographic central European area without any consideration of the international cultural influence. The aim of this research was to further investigate the management's understanding by answering the central research question based on the five sub-questions. As the first step, a study on the morals and ethics in AI and the related methodologies was conducted to create a fundamental background and an understanding of the current research and literature state (e.g. [3, 8, 10, 15–19]). Literature was imported to the analysis software ATLAS.ti, where the documents were structured and coded. In the second step, empirical research was conducted using qualitative expert interviews. Based on the central research question and the outcome of the literature study, the following main categories were defined for generating a structured interview guideline:

- Motivation: key driver for why the management started to develop an AI product or service.
- Terms and understanding: used terms and understandings of morals, ethics, and artificial intelligence within the management.
- Prioritisation: how important and how deeply anchored the three terms for the management and companies were in their daily business.
- Stakeholder plus profession: all internal and external people with their professions involved in the design and development of the AI product or service within the company.
- Guideline and manifesto: collection of points, rules, or framesets that summarise or describe how ethics could be implemented in AI systems for shaping a good future society.
- Certification: willingness and understanding of AI system certification.

- Ethical and social aspect for societies: all ethical and social functionalities, activities, or changes in AI systems that would have a strong impact on shaping future societies.
- Ethical and social aspect of their AI system: all ethical and social functionalities, activities, or changes that influences the design and development of AI systems.
- Interviewee reflection: learnings for the participant in the interview and additional input that was relevant and important to highlight from the management perspective.

The focus was to interview candidates from Austria and Germany. During the candidate recruitment process, a general AI interest was found, but people from these two countries were not motivated to openly talk about ethics. Therefore, the geographic area was broadened to also cover Scandinavian countries. The reason behind the different levels of openness to talk about ethics might be related to cultural differences, but these influences were not within the scope of this research. The recorded audio streams were processed into transcripts. These were analysed, coded with a coding frame based on main and sub-categories, shortened, and grouped on the basis of a qualitative content analysis following an approach based on Mayring [13] and Kuckartz [11]. The goal of the content analysis was to analyse the communication that was based on the recorded transcripts. The used source material (transcripts) included recorded expert interviews of executive managers (sampling unit, $n = 9$). The conducted interviews (unit of analysis) were based on the predefined and pretested interview guideline. The interview participants mainly worked as executives in different AI companies. The mix of different business industries and countries generated diversity and resulted in a collection of different perspectives and insights.

4 Management perspective

The findings of the empirical research have been summarised in the corresponding categories, and the results are presented in this section. Referring to the central research question and its sub-questions from the previous chapter, the discussion will revolve around the management considerations of ethical and moral aspects for their AI development. In the end, the motivation of the managers and their awareness of their influence on the social impact of their AI product or service will be discussed. The first sub-question of the central research topic was about the management's understanding of the used definitions and terms. Based on the outcome of the literature study, there were many existing different definitions and this question covered the perspective of the management. A common definition and understanding would make it easier to generate ethical AI guidelines and certifications at both the national and the global level. Morals were defined in different ways but represented a similar meaning to that in the existing literature. Sometimes, it was difficult for the management to find an exact distinction between morals and ethics. The term AI was often described by an example of one's own AI product or service, but as a similar approach to different literature definitions. Future AI ethics will deal with an increase of social and technical complexity and will require the perspectives of different professions. This may lead to a new diverse AI ethics discipline that will influence different existing professions, such as philosophy, psychology, law, software, and data.

The second sub-question of the central research topic was about the prioritization of moral and ethics from the management perspective. The aim was to understand if managers are setting focus on AI ethics and to understand their motivation behind the AI development. First, 77.77% of the interviewed managers ranked the prioritisation of AI ethics and moral as high. Second, the answers from the open questions identified an overlap of AI ethics, data ethics and company values. There were approaches to take the internal company values and ethical guidelines as a template to design ethics for their AI product or service. Often, ethical company principles are existing in an implicit way and they are not written in a document. This does not guarantee that every employee knows them or even understands them in a similar way, especially if the company has multi-cultural employees. Besides this, AI systems can require further ethical principles that are not used within company values. Another aspect was that companies can develop AI components that do not harm humans and their privacy, from the interviewed managers' perspective, but the final AI product or service, developed by another company, may do. In the interviews, managers answered that in some cases, they are not aware of their customers' final AI system, where their AI component is used. Several managers mentioned that their younger employees would not work for unethical projects and that this is an important factor when it comes to keep and recruit new employees. Therefore, the AI ethics topic is not only relevant for the product or service, it represents the company and its brand as well. When it comes to the motivation behind the foundation of an AI company, most of the interviewed managers mentioned customer, product (data) or curiosity as a driving force. The goal was either to solve a customer problem and to increase customer experience or to predict future state by data analytics. Both cases, customer and product, are business oriented and it is not clear how strong AI ethics is represented by the final AI product or service.

The third sub-question of the central research topic was about answering if there is an existing diversity in the AI development within the companies. Are technical IT people or other professions leading the AI development process?

The interview answers showed that the majority of internal AI development stakeholders have a technical background in computer science, data science, software engineering, mathematics, statistics and physics. The strong technical focus is not generating a diversity of different professions and also current job offerings are mainly focusing on employees with technical AI skills. Diversity is one factor to reduce bias. Further involved stakeholders are company managers, project managers, legal advisers, business developers, sales and marketing managers and customers. Employees with psychology background that does not necessarily require a university degree, are working in the field of user experience, in human interaction research projects or marketing. It was mentioned that the external stakeholder 'customer' considers AI ethics less than the AI system development company. This shows that the ethical responsibility lies on the AI system manufacturer side and, therefore, the management of the AI company needs to understand and implement ethics for their AI systems. It has earlier been mentioned that the interviewed companies lack people with social studies background and the reason behind has therefore been analysed:

- The company is too small: it is though unclear how the size of a company is defined and whether this is related to headcount, revenue or amount of sold units. Taking social responsibility shall not be a question of a company size.
- Philosophy does not match the current AI development: this can mean that philosophical aspects never change or that AI systems are shaping the future of philosophy. It requires flexibility to combine and develop AI philosophy on both sides.
- Not important enough for the business: other tasks are higher prioritised than AI ethics. Business shapes the economy and influences the social wealth, status, and societies. A reflection and strong prioritisation are required by the management.
- No consideration by the management: executives need to be aware of the importance of AI ethics for the development. If AI ethics is not implemented by manufacturers, AI systems may not operate with ethical frameworks or learn unethical behaviour.

The fourth sub-question of the central research topic was about the management's knowledge and use of any AI guidelines and principles within their company. Besides this, the managers were asked about their opinions about advantages and disadvantages of ethical AI guidelines, who should take the creator lead and their perspective about an ethical AI certification. Some managers mentioned that they have a company value document that gives a guideline to employees and management. Those guidelines have a universal character but are not explicitly referring to AI ethics. It is about a common frameset of company values that can be used as a basis for ethical AI guidelines. Most of the companies do not have any explicit ethical AI guideline. The reason for that is based on arguments, such as the AI product or service, is not violating human rights, harming people, or the management does not see any need and rated the topic as less important. What is still unclear is the ethical use of the overall AI system by the customer. The company's argument also demonstrates that actions would be taken or considered only if the AI system might harm in any way. There is no consideration to take any action as a prevention or a way to change the future development direction in a positive way, if the AI system might harm in any way, but it is not seen as a prevention or a way to change the future development direction in a positive way. During the interviews, some managers realised the benefits and the needs to create an internal ethical AI guideline for the company, as it is easy to implement into the existing business and the employees will also accept and live the principles.

Another ongoing discussion is about the core principles of ethical AI guidelines. One approach during the empirical research was to verify the results of Rothenberger et al. [17] from the management's perspective. The task for the managers was to bring six guideline proposals in order depending on their importance for them. 'Responsibility' was ranked as most important and 'Protection of Data Privacy' as second highest in both independent research studies. As the top two ranks are similar, those two are key principles for the development of general ethical AI guidelines. According to most of the managers every country shall have a representative with a good understanding in AI technology having the role as a creator of AI ethical guidelines. This includes a diverse multi-profession group of philosophers, people with religious background, AI engineers, politicians, and governmental institutions. The start of the AI guideline development should be based on law regulations but should also be flexible to be regularly updated. Many managers did not see the government in the lead due to missing experts and AI knowledge, but some said that the start of developing ethical AI guidelines could be triggered by a political action. Some could also see the European Commission, industry association, or existing ethical review committee in the lead. As there is already work ongoing for example by the European Commission, which most of the managers were not aware of, it is unclear whether the communication of those institutions does not reach the companies or if the managers set their focus in completely different directions. If all stakeholders (whole society) will not come together and align about one AI guideline, we will end up with several independently created ethical AI guidelines by different actors without any implementation. Another management input was the consideration to use a public guideline as a blueprint for customization within the own company. An

implementation can be ensured by rulesets similar to financial audits or General Data Protection Regulation (GDPR) implementations. Besides the amount and definition of the ethical AI principles, the willingness for a (global) implementation will play an important success factor. An ethical AI certification system offers transparency to others and the managers see the certification as a preventive action that can also be used as an internal review tool. Besides, it can support the commercialisation of their AI product or service and generate additional marketing value and trust. Some managers can imagine themselves certifying their AI product or service if there is a common standard and if it is easy to implement. High certification costs and that the certification is not broadly known within societies were some of the management's concern. If customers would ask for a certificate or explanation, what did not happen to them yet, most of the managers could consider an ethical AI certification. Nevertheless, the management sees the AI certification coming and this will be most likely be driven by the automotive industry leaders.

The fifth sub-question of the central research topic was about what kind of social responsibility the management has when it comes to their AI product or service. The first part included answers about ethical aspects of AI systems, environmental, social and societal impacts, ending with answers about legal system adaptions. A manager described an ethical aspect of AI systems by focusing on the goal not to maximize human happiness, but on reducing the pain for large parts of the society. Some examples that were mentioned are: investment in AI medical or agriculture applications instead of AI weapon systems. AI systems have the possibility to be unsafe, harm or discriminate people based on biased data algorithms and change the future world order as a concern in line with what research shows. Despite these results, most of the interviewed companies are still not using any ethical AI guideline for their own products or services. Software systems can only be as ethical as their developers decide to and AI/AGI systems might learn on top of it. For that reason, an early ethical AI standard will generate a long-term impact for humans, and it might become a general topic for developing software systems. Another aspect is that humans need to understand AI technology to overcome their fear or to change the picture of a Terminator robot. Therefore, people from all age groups will need to learn, interact with AI, and create their own opinions. AI developing companies, although they do not have direct end-consumer contact, are playing an important role in the whole ecosystem.

Companies of the interviewed management do not take environmental actions for their AI development based on arguments that the company is (still) too small. It would make a difference if the company would operate a high number of servers to reduce costs. The decision-making argument to switch to alternative environmentally friendly products and services was, therefore, based on costs. For the management, it is difficult to identify the right supplier, as they are using environmentally friendly marketing slogans that are not verifiable. Another argument that was mentioned is the existing dependency on main supplier for example for cloud services, where it is difficult and expensive to change to another company. Managers said that it is still too early to gain a broad awareness of the environmental impact of AI systems and they believe that companies do not care about it.

A frequent discussion topic regarding social impact of AI is the way how human will handle job loss and job changes in the future. Overall, this topic is not new and started already before automation of production processes during the industrialisation. However, the factor speed is different this time. Changes are happening faster, and humans will need to develop even more flexibility in the future. Management is aware of possible job losses, but on the other hand, there are also concerns about learning new technologies and thereafter to adapt the companies. Management is focusing on cost savings, growth, process optimization and sees the social impact of AI systems as a competitive element and not as a social problem. Arguments like the generation of new jobs and human-assistive products were mentioned. The management is not certain whether the responsibility of possible future job loss is a task of the company management or not. Another approach can be that the government decides and limits the research and development of future AI technologies to special industry sectors like for example health care.

Focusing on societal impact of AI systems, the managers pointed out how differently countries like for example Germany, USA and China handle data privacy and datasets as input data for AI algorithms. This generates use cases that are not allowed or limited in other countries and leads to that less restricted countries become fast-developing AI-nations. In contrast, a high data privacy regulation can protect human's privacy and diversity. Overall, AI systems will influence global society in many sectors like health care, retail market, elderly care, education system, and jobs. Some of the interviewed managers described their AI product or service as generating less impact on societies. Hence, the AI systems should be categorised in high and low influencers based on several criteria. The majority of the interviewed managers could not imagine the government as a leader for AI law and guidelines due to the fact that the pace of AI technology development is faster than the government's decision-making process. The managers considered the government as a high-level institution creating general framework that will be further developed and updated by a group of public institutions, work councils, AI and legal experts as representatives of the society with diverse professions. It was also proposed to include data and software developing ethics as a part of AI ethics. The quality level of input data has a

strong influence on the behaviour of AI systems. However, the input datasets cannot always be verified according the management and, hence, they have to trust these without any quality or bias verification. The generated user and machine data from the past years might also be used in future AI systems.

The interviewed managers defined transparency of AI systems as an important topic, which prevents the creation of a complex 'black box' that nobody can understand its function or behaviour. On the other hand, the managers had some concerns about being transparent when it comes to their own AI product or service. Due to competitive related issues, companies want to protect their intellectual property and are not willing to be completely open. Some managers mentioned that currently there is no demand for transparency from their customers and if there would be any demand this would come from the German market, which is more interested in decision-making processes. This may show a paradox situation, where companies do not want to share but require transparency at the same time. One of the interviewed companies found a solution combining other technologies like blockchain encrypted training of algorithms to create a complete transparent process. That could probably be an approach for other use cases. Robustness is defined differently within cultures, but overall it is seen as a prevention of harming humans. Companies are increasing robustness for their AI system by implementing industry security standards, doing penetration tests, and analysing data sets. As AI systems might control important systems and infrastructure facilities like power plants, the management awareness about putting effort on increasing product or service robustness is existing.

One approach to reduce bias in AI systems is to generate employee diversity by professions, age, gender, cultural background, etc. and to use quality input data for AI algorithm training. Since humans might also have a bias, it is difficult to generate completely non-biased AI systems. Interviewed managers were aware of that their used datasets are or might be biased. This fact shall raise concern considering the impact of all accumulated bias in the future when the complexity is increased. The future discussion would also require a classification of the bias types as not every bias is necessarily negative. Depending on the situations and the circumstances, a bias can be seen as positive by humans, for example a belief to trust most of the other people. Another mentioned aspect is to use the term fairness instead of bias. However, it is important to keep in mind that fairness is not a uniquely defined term as it is based on culture, socialization and experience and should, therefore, not be used in this context. Current AI systems are good in finding patterns, so it would be a positive use case to train and install AI algorithms to identify biased data. Privacy and (data) security is handled differently within countries and Europe might have a leading position regarding data protection. The interviewed managers from European countries do have a strong awareness about data privacy based on different regulations like the GDPR. An aspect to highlight is that an AI algorithm will understand input datasets as characters, but it will not understand whether this is a personal data from humans or not. This needs to be defined beforehand to gain people's trust in AI systems. According to the management, the key to break through digital future in big European countries is to have more people with strong technical background in computer science and data engineering among the C-level board. Regarding accountability, the management could consider different approaches, which are coming from software developing processes, on how to handle AI system failures. It is not clear if accountability of AI systems will need additional procedures since this will involve more stakeholders.

5 Conclusion

This research showed how complex and still partly unanswered the topic about the ethics of AI from a management perspective is. Besides the technological complexity, it also affects other disciplines, professions, and nations that need to cooperate, shape, and implement global frameworks and standards. Furthermore, digital technologies and trends, partly driven by AI, will have an immense impact on the consumer behaviour, economy, societies, and other sectors. Industry sectors, such as the retail market, will need to positively transform themselves into a successful combination of online and offline services. Growing unemployment rates and heavy job losses are the frequently discussed topics and a universal basic income (UBI) might be a solution in the future. Past experiences, such as the invention of the steam machine and that of electricity, show that humanity and economy have been evolved, and new ages have emerged. The development of AI might result in an increased number of software engineers and data scientists in the future. Digitalisation is a powerful transformation for global players and monopolists giving them the opportunity to further their market power. The aim of this research was to analyse the management perspective and awareness about ethics in AI. The results of the interviews revealed new perspectives and information. However, there are still many undefined and non-regulated issues to solve.

Funding Open access funding provided by University of Graz. The author acknowledges the financial support for open access publishing by the University of Graz.

Compliance with ethical standards

Conflict of interest The author declares that there is no conflict of interest.

Availability of data and material Not applicable.

Code availability Not applicable.

References

1. Anderson, M., Anderson, S.L.: Machine ethics: creating an ethical intelligent agent. AI Mag. **28**(4), 15 (2007). https://doi.org/10.1609/aimag.v28i4.2065
2. Baker-Brunnbauer, J.: Business model innovation in a paradoxical area of conflict (executive summary). https://doi.org/10.13140/RG.2.2.24272.66566 (2019). Accessed 10 Jan 2020
3. Bostrom, N., Yudkowsky, E.: The ethics of artificial intelligence. In: Frankish, K., Ramsey, W. (eds.) The Cambridge handbook of artificial intelligence, pp. 316–334. Cambridge University Press, Cambridge (2014). https://doi.org/10.1017/CBO9781139046855.020
4. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners. https://arxiv.org/abs/2005.14165 (2020). Accessed 9 June 2020
5. Continental: ethics regulations for artificial intelligence. https://www.continental.com/en/sustainability/news/news-2020/ai-code-of-ethics-224686 (2020). Accessed 9 June 2020
6. EC: ethics guidelines for trustworthy AI. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419 (2019). Accessed 31 Mar 2020
7. EC: white paper on artificial intelligence: a European approach to excellence and trust. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (2020). Accessed 19 Feb 2020
8. 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.: AI4People—an ethical framework for a good AI society: opportunities risks, principles, and recommendations. Minds Mach. **28**, 689–707 (2018). https://doi.org/10.1007/s11023-018-9482-5
9. FRA: the European Union Agency for fundamental rights: AI policy initiatives (2016–2020). https://fra.europa.eu/en/project/2018/artificial-intelligence-big-data-and-fundamental-rights/ai-policy-initiatives (2020). Accessed 4 July 2020
10. Indurkhya, B.: Is morality the last frontier for machines? New Ideas Psychol. **54**, 107–111 (2019). https://doi.org/10.1016/j.newideapsych.2018.12.001
11. Kuckartz, U.: Qualitative Inhaltsanalyse. Methoden, Praxis, Computer-unterstützung. Beltz, Weinheim (2018)
12. Matzler, K., Bailom, F., Friedrich von den Eichen, S., Anschober, M.: Wie Sie Ihr Unternehmen digital auf das digitale Zeitalter Disruption vorbereiten. Vahlen, Munich (2016)
13. Mayring, P.: Qualitative Inhaltsanalyse: Grundlagen und Techniken. Beltz, Weinheim (2015)
14. Moor, J.H.: The nature, importance, and difficulty of machine ethics. IEEE Intell. Syst. **21**(4), 18–21 (2006). https://doi.org/10.1109/MIS.2006.80
15. Pagallo, U.: When morals ain't enough: robots, ethics, and the rules of the law. Mind Mach. **27**, 625–638 (2017). https://doi.org/10.1007/s11023-017-9418-5
16. Ransbotham, S., Khodabandeh, S., Fehling, R., LaFountain, B., Krion, D.: Winning with AI, MIT sloan management review and boston consulting group. https://sloanreview.mit.edu/ai2019 (2019). Accessed 28 Nov 2019
17. Rothenberger, L., Fabian, B., Arunov, E.: Eelevance of ethical guidelines for artificialintelligence – a survey and evaluation. In: Proceedings of the 27th European conference on information systems (ECIS), Stockholm and Uppsala, Sweden, https://aisel.aisnet.org/ecis2019_rip/26 (2019). Accessed 29 Oct 2019
18. Siau, K., Wang, W.: Ethical and moral issues with AI - a case study on healthcare robots, emergent research forum (ERF). https://www.researchgate.net/publication/325934375_Ethical_and_Moral_Issues_with_AI (2018). Accessed 20 Oct 2019
19. Tegmark, M.: Life 3.0. Penguin Random House UK, London (2018)
20. TÜV-Verband Thüringen: Verbraucher wollen Sicherheit und Transparenz bei Künstlicher Intelligenz. https://www.tuev-thueringen.de/unternehmen/presse/texte/artikel/tuev-verband-verbraucher-wollen-sicherheit-und-transparenz-bei-kuenstlicher-intelligenz (2020). Accessed 31 Mar 2020
21. Wachter, S., Mittelstadt, B. & Russell, C.: Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI. https://doi.org/10.2139/ssrn.3547922 (2020). Accessed 10 June 2020