5-15-2019

RELEVANCE OF ETHICAL GUIDELINES FOR ARTIFICIAL INTELLIGENCE – A SURVEY AND EVALUATION

Lea Rothenberger
Hochschule für Telekommunikation Leipzig (HfTL), lrothenberger@t-online.de

Benjamin Fabian
Humboldt University Berlin, bfabian@wiwi.hu-berlin.de

Elmar Arunov
Telekom Innovation Laboratories, elmar.arunov@telekom.de

RELEVANCE OF ETHICAL GUIDELINES FOR ARTIFICIAL INTELLIGENCE – A SURVEY AND EVALUATION

Research in Progress

Rothenberger, Lea, University of Telecommunications Leipzig (HfTL), Leipzig, Germany, lrothenberger@t-online.de

Fabian, Benjamin, Humboldt University Berlin, Berlin, Germany, & University of Telecommunications Leipzig (HfTL), Leipzig, Germany, bfabian@wiwi.hu-berlin.de

Arunov, Elmar, Telekom Innovation Laboratories, Berlin, Germany, elmar.arunov@telekom.de

Abstract

Ethics for artificial intelligence (AI) is a topic of growing practical relevance. Many people seem to believe that AI could render jobs obsolete in the future. Others wonder who is in charge for the actions of AI systems they encounter. Providing and prioritizing ethical guidelines for AI is therefore an important measure for providing safeguards and increasing the acceptance of this technology.

The aim of this research is to survey ethical guidelines for the handling of AI in the ICT industry and evaluate them with respect to their relevance. For this goal, first, an overview of AI ethics is derived from the literature, with a focus on classical Western ethical theories. From this, a candidate set of important ethical guidelines is developed. Then, qualitative interviews with experts are conducted for in-depth feedback and ranking of these guidelines. Furthermore, an online survey is performed in order to more representatively weight the ethical guidelines in terms of importance among a broader audience. Combining both studies, a prioritization matrix is created using the weights from the experts and the survey participants in order to synthesize their votes. Based on this, a ranked catalogue of ethical guidelines for AI is created, and novel avenues for research on AI ethics are presented.

Keywords: Artificial Intelligence, Ethics, Ethical Guidelines.

1 Introduction

The issue of responsibility is just one of many ethical challenges that arise during the use of systems that are based on artificial intelligence (AI). The aim of this research is to develop a solution: An evaluated catalogue of important guidelines to mitigate the emergence of such ethical problems. Such an evaluation should reflect both the opinions expertise of AI experts but also of a broader audience of people. This could support all stages of the design, development, deployment and use of AI systems in modern society.

1.1 Theoretical Foundations and Motivation

Artificial Intelligence is usually understood as a subfield of Cognitive Computing (Sommer, 2017). The term "Artificial Intelligence" itself embraces two aspects. Intelligence can be understood as a set of higher cognitive abilities that apparently (and from a classical point of view) sets human beings apart from other animals. John McCarthy, a pioneer of the field, was the first to define AI in the year 1955 as "the science and engineering of making intelligent machines" (McCarthy, et al., 1955). A crucial question is how to measure intelligence, which also becomes important for understanding AI.

On the other hand, the term "artificial" may also raise negative connotations among the general human population. It may be associated with fears, such as those instilled by robots and erratically behaving

autonomous computers from Science Fiction literature, movies, or games (Gunkel, et al., 2012). Such connotations, among other reasons such as using AI systems for warfare or general surveillance, emphasize the general importance of discussing ethical guidelines to the field of AI.

In the field of ethics, we focus on Western classical ethic theories at the current stage of our research. An important aspect of future research is how common and global ethical guidelines can be derived in accordance with intercultural computer and information ethics (Capurro, 2008) (Britz & Hongladarom, 2010) (Mingers & Walsham, 2010) (Bynum, 2001) (Mittelstadt, et al., 2016) and ethical pluralism (Ess, 2006), which is in itself an important field of discussion in ethical research (Hiruta, 2006) (Bynum, 2001).

In classical Western ethics, there are several frameworks for accurately describing and reasoning about ethical attitudes and responsibility, including the important field of teleology and deontology. The term "**τέλος**" from Greek can be translated as "end" or "goal": In teleology, a norm or action is considered as good if it leads to good consequences. Important teleological approaches include utilitarianism and its predecessor from antiquity, hedonism.
In contrast, deontology does not emphasize the consequences of an action, but underlying values and principles. It focuses on "**δέον**", i.e., that which is necessary, right or duty. From this point of view, the ethically good is based on the right principles, virtues and methods. Here, expressed with some simplification, not the good action determines the right one, but the right action determines the good one. One of the most important deontological approaches is called virtue ethics (Vallentyne, 1987).

One of the first authors who discussed guidelines or laws for robots and intelligence systems was Isaac Asimov. In his short story "Runaround" from 1942 (later also published in the book "I Robot"), he created the well-known *Three Laws for Robotics* (Asimov, 2004), which inspired many researchers in the field and became the foundation for all later ethical guidelines for designing and using AI:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given 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 First or Second Law. (Asimov, 2004)

Decades of research in the field of AI have moved it from science fiction to real-world implementations and human-robot interactions (Sheridan, 2016), sometimes met with positive attitudes (Ray, et al., 2008), but sometimes also instilling concerns and fears among the human population, including AI experts (Bostrom, 2014) (Bohannon & Russell, 2015) (Gent, 2015) (Johnson & Verdicchio, 2017) (Cave & Dihal, 2019). In fact, industrial robots have already been the cause of injuries and even death (Kirschgens, et al., 2018). Current and future AI systems are no longer confined to controlled industrial environments where they are interacting only with trained workers. Companies are directing considerable attention to developing service robots that will perform basic household tasks or assist elderly people at home. Robot pets, vacuum cleaners, transport systems in hospitals or robot guides in museums are becoming commonplace (Allen, et al., 2006).

These robots are becoming closer to the human; this also means that they have their own moral status, i.e., counting as a moral entity and be permissible to do things from its own will (Bostrom & Yudkowsky, 2014). But is it the right approach to put the moral status above the will of a human? For example, if an elder care robot has the task to take care of the human living very healthily, but the biggest or even last wish of the human is to eat very unhealthy food and live his life like he wants. Here the question is: what is morally 'right' (Burton, et al., 2017)? These questions have stimulated the research field of *Machine Ethics* (Moor, 2006) (Allen, et al., 2006) (Anderson & Leigh Anderson, 2011) (Allen & Wallach, 2009) (Anderson & Leigh Anderson, 2007). The need for ethical guidelines for (big) data analytics (Someh, et al., 2016) (Saltz & Dewar, 2019), machine learning and AI is further substantiated by current research literature (Bryson, 2017) (Etzioni & Etzioni, 2017) (Lesser, et al., 2018).

There are further ethical dilemmas in recent practice, for example, when a self-driving car has to make a decision if it is better to save the lives of the three people outside the car instead of the one life of the driver (Goodall, 2014) (Bonnefon, et al., 2016). Are there pre-defined sets of values, or on which basis the car has to make the decision (Pavaloiu & Köse, 2017)? One of the most significant factors is human bias. When the car needs to make the decision on its own, there could also be bias existing in the algorithm of the car, triggered by a human factor (Wang & Siau, 2018). The most frequent biases include gender bias (Larson, 2017), which means having disadvantages because of the gender, and race bias (Koolen & van Cranenburgh, 2017), i.e., disadvantages because of race.

Further concerns refer to social aspects such as AI replacing human beings in their jobs (Arntz, et al., 2016) (McClure, 2018) (Acemoglu & Restrepo, 2018). In robotic process automation, for example, it is estimated that one in five job positions will become obsolete within the next ten years (Horváth & Partner, 2018). On the other hand, around two-thirds of respondents of another study believe that AI will support society with solving complex problems, but may also replace certain professions (Bothun, Lieberman, & Rao, 2017).

To sum up, there are three areas of ethical and social concerns: Safety and errors, law and ethics and the social impact, where most challenges arise (Lin, et al., 2011). In our research, we aim at bridging the gap from ethics to technical design, implementation and operation by surveying and evaluating ethic guidelines for AI.

1.2 Methods

In this explorative study, we first conduct a literature survey and an overview of existing ethical guidelines for AI and corresponding research from industry, academia, governments and other institutions. The relevant findings are summarized and discussed. From this, the most prevalent ethical guidelines across all sources are identified and collected for further ranking of their importance by experts via qualitative interviews and also an online questionnaire aimed at the general population. For this, supporting definitions and explanations are added to each guideline to facilitate a common understanding. For the qualitative interviews, leading experts are identified and contacted, both internally from a global ICT corporation, as well externally. Seven extended interviews are conducted and systematically analysed using the Grounded Theory methodology. Based on these interviews, a specific question set is created for the quantitative online questionnaire with 51 respondents belonging to a broader audience. The results from both studies are analysed individually, but are also combined for synthesis with en equal weighting of ranking votes from interviews and online questionnaire.

2 Survey of Guidelines for Ethical Handling of AI

2.1 Overview of Guidelines for AI Ethics

In our overview, the investigated parties are classified according to the type of organization (industry, academia, research and development institution, government and association), and the corresponding guidelines or focus fields are summarized. For space reasons, the shortened version given in Table 1 only compares three industrial parties and one party each from academia, research and development institutions, governments and associations.

Name of Organization	Type of Organization	Focus Field/ Guideline
Microsoft	Industry	Microsoft AI Principles: • Fairness: AI should maximize efficiency without destroying dignity and protect against bias. • Responsibility: An AI has algorithmic responsibility. • Transparency: An AI should be transparent. • Ethics: An AI should have humaneness and has been designed

		for intelligent privacy. (Microsoft, 2018)
Yale University	Academia	The field of machine morale is driven by a series of questions: • Do we need robots that are capable of making moral decisions? When and for what? • Do we want the computer to make moral decisions? • Are robots the kind of beings who are capable of making moral judgements? • Whose morale or morality should be implemented? • How can we make ethics computer-compatible? (Wallach, 2010)
AI2 (Allen Institute for Artificial Intelligence)	Research and Development Institution	• An AI system must be subjective to the full range of laws that apply to its human operator. • An AI system must clearly disclose that it is not human. • An AI system can not back up or disclose confidential information without the explicit permission of the information owner. (AI-Ethics, 2017)
German Federal Government	Government	Strategy Paper for AI: • Inviolability of human dignity • Respect for privacy • Principle of equality • Safety, efficiency, sustainability • Form a Data Ethics Commission (Bundesregierung, 2018)
PAI (Partnership on AI)	Association	• Objectives: To develop and share best practices, to broaden public understanding, to provide an open and inclusive platform for discussion and engagement, to identify and promote AI's intended social benefits • Working groups: safety-critical AI, fairness, transparent and responsible AI, cooperation between people and AI, AI lab and business, social and social impact on AI, AI for specific initiatives (Burgess, 2017)

Table 1. Overview of AI Ethics (Shortened Version)

This overview sets out a proposal for the catalog of guidelines, which will be examined in the following sections for relevance. The most common guidelines were extracted as shown in the following Table 2, which also provides the type of organizations issuing the benchmark and a definition for each guideline.

Guideline	**Type of Organization**	**Definition**	**Extract of Source Guideline from Benchmark**
Transparency	Government, Association, Research and Development institution	An AI system must be transparent about being an AI (before usage or interaction). Therefore, an international standard has to be launched.	Human identity in the age of AI (Kurz, 2018) An AI should be transparent. (Winfield, 2018) An AI system has to reveal that it is not a human. (AI-Ethics, 2017)

Responsibility	Association	The operator and user of an AI are bearing the blame for any action (and their consequences) of the AI system.	Are robots the kind of beings that are able to make moral decisions? (Wallach, 2010) Responsible AI (Burgess, 2017) Operator of an AI should have the right authority. (Winfield, 2018)
Protection of Data Privacy	Government, Industry	Unauthorized interceptions should be avoided. The user has to agree explicitly to the usage of his private data.	Respect for privacy (Bundesregierung, 2018) Composition of ethical and judicial guidelines for the development of AI on basis of the constitutional law (Commission, European, 2018) Ethics: An AI should have humaneness and has been designed for intelligent privacy. (Microsoft, 2018)
Bias should be minimized	Industry	Unfair, racist bias should be minimized.	Fairness: AI should maximize efficiency without destroying dignity and protect against bias. (Microsoft, 2018) Avoidance of creation or reinforcement of unfair bias (Nelson, 2018) Avoidance of bias (Accenture, 2018)
An AI should have a purpose	Industry, Association	Supporting the human should be the highest purpose of an AI. The AI must not replace the human. A human-machine cooperation model will be established.	Foster the cooperative model of human and AI. (Deutsche Telekom AG, 2018) An AI should serve the human and its environment. (Winfield, 2018)
Robustness	Academia, Association, Industry	AI algorithms should be robust against manipulations, both internally and externally. For example, a language assistant should not order any items through any external influences.	AI algorithms are robust against manipulation. (Bostrom & Yudkowsky, 2014) Cybersecurity: The need for strong protection against hacking will increase as AI systems take a heightened role in society. (Accenture, 2018)

Table 2. Selected Ethical Guidelines

2.2 Expert Interviews Using Grounded Theory

For evaluating the guidelines and their relevance, qualitative interviews with experts were conducted. The interviews were analysed with help of the *Grounded Theory Methodology (GTM)*, which is used in a slightly modified manner. After the interviews were transliterated, the first phase of *GTM*, open coding for analysing the central statements of the interview, was conducted. This was followed by the second phase, axial coding for summarizing similar codes from the first phase. The third phase of the methodology, selective coding, was not applied in the current study, because this study aimed to cover a wide range of answers. Therefore, the types of the last phase are constituted of the code paradigms from the second phase. Phenomena with the same meanings are summarized to types. The appropriate categories within the code paradigm, which represent the relations, are also summarized and provide the type dimensions. Overall, four type paradigms could be created out of seven interviews conducted so far. Figure 1 shows an example paradigm, the Compliance View (Type 4).

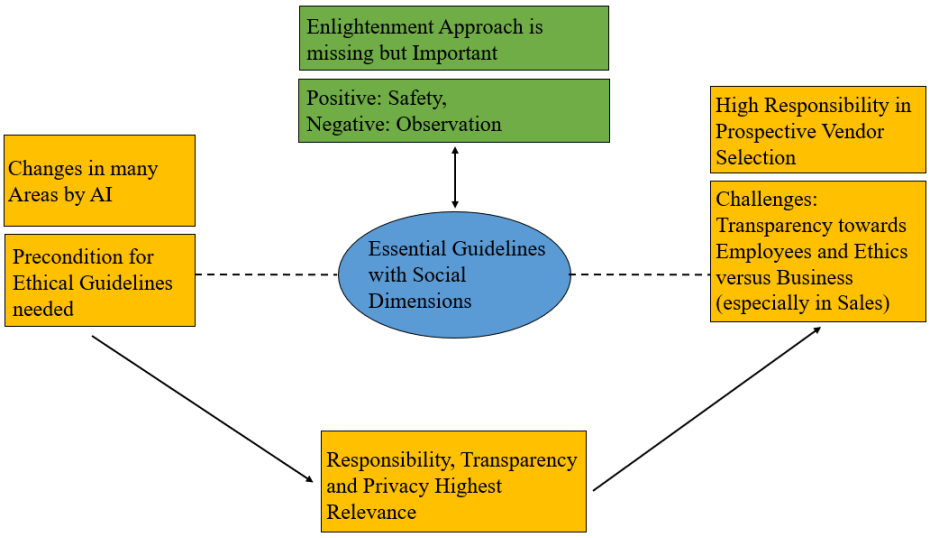

Figure 1. Type Paradigm 4 - Compliance View

The Engineering View (Type 1) and the Research View (Type 2) are not very enthusiastic about launching ethical guidelines for an AI. For the most part, they offer the opinion that sufficient scientific approaches would already exist. Furthermore, the ethical discussion in Germany would be arising only from a translation error of the term "Artificial Intelligence" into the German language, because people would be always comparing machine intelligence to human intelligence.

In contrast, the Innovation View (Type 3) and the Compliance View (Type 4) state that an ethical system of rules is strongly needed. Such guidelines would be essential and necessary because every area of life, and not only one single branch, would be changed through AI in the future. All views and types are undecided if a machine could have its own awareness.

The experts ranked the guidelines as shown in Table 3.

Guidelines	Arithmetic Mean	Rank Order
Responsibility	4.71	1
Transparency	4.43	2
Protection of Data Privacy	4.43	2
Robustness	4.14	3
Bias should be minimized	3.67	4
An AI should have a purpose	3.29	5

Table 3. Ranking of Ethical Guidelines by the Experts

2.3 Online Survey

The goal of the online survey was to provide a broader evaluation of the ethical guidelines in society as a whole, beyond the expertise of the experts. A corresponding section of the expert interview was also used as a basis for the online survey. Again, a 5-point Likert scale was applied to measure the subjective importance of each guideline according to the survey participants. Invitations to the online survey were sent with help of personal social networks on Facebook and LinkedIn, to invite many people from a broader audience for diversified opinion and evaluation of the guidelines.

Overall, 51 participants fully completed the questionnaire and are included in our data analysis. Table 4 shows the results. Nearly all evaluations of the guidelines resulted in average values of four and five, indicating a strong support for their inclusion and importance. A slight exception was the guideline "Bias should be minimized". This may be explained by several causes, for instance, the guideline might be still somewhat unclear to participants, even though exact supporting definitions were given;

or the participants did not want to comment on a guideline. The term "bias" may also involve the highest degree of previous AI knowledge among all terms in the guidelines.

Guideline	Arithmetic Mean	Rank Order
Protection of Data Privacy	4.37	1
Robustness	4.35	2
Responsibility	4.31	3
Transparency	4.12	4
Bias should be minimized	4.06	5
An AI should have a purpose	3.86	6

Table 4. Ranking of Ethical Guidelines by Participants of the Online Survey

3 Combined Results

Based on the knowledge gathered from the qualitative expert interviews and the quantitative online survey, the final ranked catalogue for ethical guidelines and their importance was created (Table 5).
A new arithmetic average is calculated with help of the rankings from the expert interviews and the online survey applying equal weights to the results from both answer sets. This combined arithmetic average is calculated using the following formula:

$$\emptyset_{new} = (\emptyset_{Experts} \times 0{,}5) + (\emptyset_{Survey} \times 0{,}5)$$

The results are rounded to two decimal places for better readability. After sorting, the outcomes are shown in Table 5, presenting the combined ranking of the experts and the online survey.

Guideline	Arithmetic Mean	Final Rank Order
Responsibility	4.51	1
Protection of Data Privacy	4.40	2
Transparency	4.28	3
Robustness	4.25	4
Bias should be minimized	3.87	5
An AI should have a purpose	3.58	6

Table 5. Combined Ranking of Ethical Guidelines

Overall, *Responsibility* has received the highest votes for relevance and hence achieved the highest rank. For both groups, the interviewed experts and the participants of the online survey, it is most important that there is someone who is declared as responsible for the actions and especially the consequences of an AI system. One expert even postulated this guideline as residing above all others and influencing all other guidelines. At the second rank, there is the guideline *Protection of Data Privacy*. Also for the participants of the online survey it is important who protects their data and what is happening with this data. Even though one expert stated that this guideline is already regulated within the constitutional law, implementing privacy-protection with AI remains an important future challenge.
Transparency is ranked at the third place. For the interviewees, there are situations where it is necessary to know that one is interacting with an AI. The guideline *Robustness* is following at the fourth rank; attackers should not be able to manipulate the algorithm from outside or to steal any data from the storage of an AI.
The second to last place was assigned to the guideline *Bias should be minimized*. In opinion of the experts, bias should not only be minimized, it should be completely avoided. On the other hand, experts were also stating that bias would only exist because of the humans training the AI system. The lowest relevance was assigned to *An AI should have a purpose*. Here, some experts offered the opinion

that an AI would not always have the purpose to support or replace humans; some robots, for example, are working in environments that are hostile to human beings.

4 Conclusion and Future Work

Our study presents a survey and evaluation of guidelines for AI ethics, including an initial ranking according to experts based on qualitative interviews and an online questionnaire among a broader audience. A clear limitation of our current study is the number of expert interview partners (n_e=7) for the qualitative and in-depth interviews; we argue that this is, in part, mitigated by including several distinguished experts from the AI domain. Nonetheless, extending the number of interviews needs to be addressed in future work. The number of participants in the online survey (n_s=51) should also be increased; then, the impact of job roles, gender, country and psychological traits could also be studied in-depth. Moreover, further guidelines could be included in the ranking, and extension to a scenario-driven motivation could lead to interesting comparisons across applications.

According to our results, responsibility was clearly ranked first; experts as well as participants of the online survey emphasized its very high relevance. However, many respondents asked who would be responsible for the actions of an AI: The AI manufacturers – because of their knowledge on potential application areas? Or the developer or programmer of an AI – because she develops or adopts the algorithms, implements them into the machine and would be responsible to protect it from manipulation? Or would the user of an AI be responsible since he could be expected to acquire some knowledge on the AI's inner workings or external behaviour and protect it from manipulation?

Furthermore, privacy protection was deemed very important and was ranked second overall. Addressing privacy in future AI systems will involve many challenges, including general trade-offs with accuracy and utility, and will also need further application-specific analyses. There was also a challenge at the evaluation of the guidelines, especially for the guideline *Bias should be minimized*. At this guideline, the concept of "bias" may also involve the highest degree of previous AI knowledge among all terms in the guidelines.

Moreover, it would be of high interest to conduct further research on ethics and address the question if ethics for human developers and AI users are needed, or ethics for machines, which also seemed to be a point of disagreement among the experts. Some respondents (mostly engineers) even stated that no ethics at all would be needed, but compliance to relevant existing engineering guidelines. This may be a reappearance of the contrasts between deontology and teleology, from a modern and technical perspective. Future research should critically address this issue and analyse if corresponding methods from engineering would be able to address the ethical challenges of current and future AI technology.

Another important aspect of future research is how common and global ethical guidelines for AI can be derived in accordance with intercultural computer and information ethics and ethical pluralism. In our current work, only the classical Western ethical perspective could be reflected.

Future research should also further reflect the different application fields of AI. Will it be necessary to differentiate ethic guidelines according to different requirements in particular application domains? Some experts argued in this direction because most (if not any) AI would be specific and tailored to an application domain. According to this view, there would be no need for a general catalogue of ethic guidelines, but for several application-specific catalogues. In a variant, there could be a "base catalogue" of guidelines that need to be implemented and enforced by any AI application, which could be extended by application-specific catalogues or different weightings of guidelines tailored to individual domains.

In summary, the field of AI ethics is still in an emerging state with strong and growing practical relevance. We hope that our current paper can stimulate the important and timely debate on AI ethics and will support future developments of ethical guidelines that directly influence AI engineering and operations.

References

Accenture. (2018). *Responsible AI and Robotics*. Retrieved August 30, 2018, from https://www.accenture.com/gb-en/company-responsible-ai-robotics

Acemoglu, D., & Restrepo, P. (2018, January). *Artificial Intelligence, Automation and Work. NBER Working Paper No. 24196.* Retrieved March 23, 2019, from https://www.nber.org/papers/w24196

AI-Ethics. (2017). *AI-Ethics - Law, Technology and Social Values*. Retrieved August 30, 2018, from https://ai-ethics.com/past-principles/

Allen, C., Smit, I., & Wallach, W. (2006). Why Machine Ethics? *IEEE Intelligent Systems, 21* (4), 12–17.

Allen, C., & Wallach, W. (2009). *Moral Machines: Teaching Robots Right from Wrong.* New York: Oxford University Press.

Anderson, M., & Leigh Anderson, S. (2011). *Machine Ethics.* Cambridge University Press. Cambridge.

Anderson, M., & Leigh Anderson, S. (2007). Machine Ethics: Creating an Ethical Intelligent Agent. *AI Magazine, 28* (4).

Arntz, M., Gregory, T., & Zierahn, U. (2016). *The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis.* Paris: OECD Publishing.

Asimov, I. (2004). *I, Robot* (The Robot Series). New York: Random House.

Bohannon, J., & Russell, S. (2015). Fears of An AI Pioneer. *Science, 349* (6245), 252.

Bonnefon, J.-F., Rahwan, I., & Shariff, A. (2016). The Social Dilemma of Autonomous Vehicles. *Science, 352*, 1573–1576.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies.* Oxford: Oxford University Press.

Bostrom, N., & Yudkowsky, E. (2014). *The Ethics of Artificial Intelligence.* New York: Cambridge University Press.

Bothun, D., Lieberman, M., & Rao, D. A. (2017). *Bot.Me: A Revolutionary Partnership.* Consumer Intelligence Series. New York: PwC.

Britz, J., & Hongladarom, S. (2010, October). *International Review of Information Ethics (IRIE).* Retrieved March 23, 2019, from http://www.i-r-i-e.net/inhalt/013/013-full.pdf

Bryson, J. (2017). Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems. *IEEE Computer, 50* (5), 116.

Bundesregierung. (2018). *Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz.* Berlin: Deutsche Bundesregierung.

Burgess, M. (2017). *Apple's head of Siri is joining the Partnership on AI – The Silicon Valley giants hope to bring more transparency to AI around questions of bias and ethics*. Retrieved August 30, 2018 (Updated January 27th, 2019), from http://www.wired.co.uk/article/ai-partnership-facebook-google-deepmind

Burton, E., Goldsmith, J., & Walsh, T. (2017). Ethical Consideration in Artificial Intelligence Courses. *AI Magazine, 38* (2).

Bynum, T. (2001, August 14). *Computer and Information Ethics.* In: *Stanford Encyclopedia of Philosophy.* Retrieved March 23, 2019, from http://plato.stanford.edu/entries/ethics-computer/

Capurro, R. (2008). Intercultural Information Ethics: Foundations and Applications. *Journal of Information, Communication and Ethics in Society, 6* (2), 116-126.

Cave, S., & Dihal, K. (2019). Hopes and Fears for Intelligent Machines in Fiction and Reality. *Nature Machine Intelligence, 1*, 74-78.

Commission, European. (2018). *Call for A High-Level Expert Group on Artificial Intelligence*. Retrieved August 30, 2018, from https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence

Deutsche Telekom AG. (2018). *Deutsche Telekom's guidelines for artificial intelligence*. Retrieved October 8, 2018, from https://www.telekom.com/en/company/digital-responsibility/details/artificial-intelligence-ai-guideline-524366

Ertel, W. (2016). *Grundkurs Künstliche Intelligenz - Eine praxisorientierte Einführung* (4. Auflage). Wiesbaden: Springer Vieweg.

Ess, C. (2006). Ethical Pluralism and Global Information Ethics. *Ethics and Information Technology, 8* (4), 215–226.

Etzioni, A., & Etzioni, O. (2017). Incorporating Ethics into Artificial Intelligence. *The Journal of Ethics, 21* (4), 403–418.

Gent, E. (2015). AI: Fears of 'Playing God'. *Engineering & Technology , 10* (2), 76–79.

Goodall, N. J. (2014). Machine Ethics and Automated Vehicles. *Road Vehicle Automation. Lecture Notes in Mobility*, 93–102.

Gunkel, D. J., Byson, J. J., & Torrance, S. (2012). Behind the Mask: Machine Morality. In K. Miller, M. J. Wolf, & F. Grodzinsky (Eds.), *The Machine Question: AI, Ethics and Moral Responsibility*. Birmingham: AISB/IACAP World Congress 2012, 34–38.

Hiruta, K. (2006). What Pluralism, Why Pluralism, and How? A Response to Charles Ess. *Ethics and Information Technology, 8* (4), 227–236.

Horváth & Partner. (2018). *Automation: jede zehnte Stelle in fünf Jahren obsolet?* Retrieved October 5, 2018, from https://www.consulting.de/job-karriere/10-fragen-an/consulting/automation-jede-zehnte-stelle-in-fuenf-jahren-obsolet/

IZW, Impulszentrum Zukunftsfähiges Wirtschaften. (2018). Ethik, Moral und Verantwortung. In *Wirtschaftsethik, Kritisch betrachtet*. Graz: Impulszentrum Zukunftsfähiges Wirtschaften (IZW), 4-26.

Johnson, D. G., & Verdicchio, M. (2017). AI Anxiety. *Journal of the Association for Information Science and Technology, 68* (9), 2267–2270.

Kirschgens, L. A., Ugarte, I. Z., Uriarte, E. G., Rosas, A. M., & Vilches, V. M. (2018). *Robot Hazards: From Safety to Security.* Alava: Alias Robotics S.L.

Koolen, C., & van Cranenburgh, A. (2017). *These Are Not the Stereotypes You Are Looking For: Bias and Fairness in Authorial Gender Attribution.* Valencia: Association for Computational Linguistics.

Kurz, C. (2018). *Ethische Fragen bei Künstlicher Intelligenz: Mit welchen Herausforderungen müssen wir umgehen?* Retrieved August 30, 2018, from https://netzpolitik.org/2018/ethische-fragen-bei-kuenstlicher-intelligenz-mit-welchen-herausforderungen-muessen-wir-umgehen/

Larson, B. N. (2017). *Gender As A Variable in Natural-Language Processing: Ethical Considerations.* Valencia: Association for Computational Linguistics.

Lesser, V. R., Leung, C., Miao, C., Shen, Z., Yang, Q., & Yu, H. (2018). Building Ethics into Artificial Intelligence. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*.

Lin, P., Abney, K., & Bekey, G. (2011). Robot Ethics: Mapping the Issues for A Mechanized World. *Artificial Intelligence, 175*, 942–949.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.

McClure, P. K. (2018). "You're Fired," Says the Robot: The Rise of Automation in the Workplace, Technophobes, and Fears of Unemployment. *Social Science Computer Review, 36* (2), 139–156.

Microsoft. (2018). *Microsoft AI Principles*. Retrieved August 30, 2018, from https://www.microsoft.com/en-us/ai/our-approach-to-ai

Mingers, J., & Walsham, G. (2010). Toward Ethical Information Systems: The Contribution of Discourse Ethics. *MIS Quarterly, 34* (4), 833–854.

Mittelstadt, B. D., Stahl, B. C., & Timmermans, J. (2016). The Ethics of Computing: A Survey of the Computing-Oriented Literature. *ACM Computing Surveys (CSUR), 48* (4), 55.

Moor, J. H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems, 21* (4), 18–21.

Nelson, D. (2018). *Google's DeepMind Launches AI Ethics Research Unit*. Retrieved August 30, 2018, from https://sciencetrends.com/googles-deepmind-launches-ai-ethics-research-unit/

Pavaloiu, A., & Köse, U. (2017). Ethical Artificial Intelligence - An Open Question. *Journal of Multidisciplinary Developments, 2* (2), 15–27.

Ray, C., Mondada, F., & Siegwart, R. (2008). What Do People Expect From Robots? *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems.*

Saltz, J. S., & Dewar, N. (2019). Data Science Ethical Considerations: A Systematic Literature Review and Proposed Project Framework. *Ethics and Information Technology*, 1–12.

Sheridan, T. B. (2016). Human–Robot Interaction: Status and Challenges. *Human Factors - The Journal of the Human Factors and Ergonomics Society, 58* (4), 525–532.

Someh, I. A., Breidbach, C. F., Davern, M. J., & Shanks, G. G. (2016). Ethical Implications of Big Data Analytics. *Proceedings of the European Conference on Information Systems (ECIS)*.

Sommer, P. (2017). *IBM Digital Nordic*. Retrieved November 26, 2018 from https://www.ibm.com/blogs/nordic-msp/artificial-intelligence-machine-learning-cognitive-computing/

Vallentyne, P. (1987). The Teleological/Deontological Distinction. *Journal of Value Inquiry*, 21:1.

Wallach, W. (2010). Robot Morals and Human Ethics: The Seminar. *11* (1), 85–90.

Wang, W., & Siau, K. (2018). *Ethical and Moral Issues with AI - A Case Study on Healthcare Robots.* Missouri: Emergent Research Forum (ERF).

Winfield, A. (2018). *A round up of robotics and AI ethics: part 1 principles*. Retrieved August 30, 2018, from https://robohub.org/a-round-up-of-robotics-and-ai-ethics-part-1-principles/