# The Current State of Industrial Practice in Artificial Intelligence Ethics

Ville Vakkuri, Kai-Kristian Kemell, Joni Kultanen, and Pekka Abrahamsson

**Abstract**—As Artificial Intelligence (AI) systems become increasingly widespread, we have begun to witness various failures highlighting issues in these systems. These incidents have sparked public discussion related to AI ethics and further accelerated the on-going academic discussion in the area. High-level guidelines and tools for managing AI ethics have been introduced to help industry organizations make more ethical AI systems, but we currently know little about the state of industrial practice. Have these guidelines been adopted by the software industry for developing AI solutions? Are these failures that make the news just the tip of the iceberg? We provide insights into the current state of practice by presenting the results of a survey of 211 software companies.

**Index Terms**—Artificial Intelligence, Ethics, Software Development

## 1 Introduction

VARIOUS technologies related to Artificial Intelligence (AI) have been at the top of the Gartner Hype Cycle for Emerging Technologies for years. Organizations from across industries are looking for ways to reap benefits from utilizing AI in different ways. During our recent visit to Slush, the world's leading startup and tech event with 25 000 attendees, we saw the booths of the AI startups downright flooded, with lines forming on occasion. In general, the hype surrounding AI has long since reached a fever pitch.

As AI technologies become increasingly widespread, they start to exert a society-wide influence. Most of us interact with AI systems every day as consumers and customers, mostly without even realizing it. As the number of AI systems grows, so does the number of AI system failures we witness.

Various high-profile incidents that have made the global news have sparked public discussion on AI ethics. A growing number of voices, both from researchers and media, as well as governments, have called for more ethical AI systems in the wake of these failures. Sometimes these incidents are a result of simply not knowing better, as was the case with the Amazon recruitment AI that became biased against women [15]. Having been trained using past recruitment data, the AI saw mostly men hired, and learned thus that they were preferable hires.

On the other hand, sometimes these incidents are simply about intentional misconduct. While it was more of a lesson in relation to data handling in general, the case of Cambridge Analytica is one such example. Cambridge Analytica utilized data from the users of Facebook without their consent to use for political advertising purposes [18]. Even though they were not the ones misusing the data themselves, it resulted in Facebook taking a publicity hit as well. With AI systems typically handling vast amounts of data, questions of data governance are important. The temptation to gather any and all data that may or may not be useful one day can be high when dealing with AI.

Yet, despite all the talk in the area recently, outside these incidents highlighting failures, we know little of the current state of practice of ethics in AI. Software engineering researchers have recently begun to understand more broadly how artificial intelligence and machine learning are changing the way the software is being developed [13]. Has the public and academic discussion in the area motivated smaller industry players to develop more ethical AI?

To the best of our knowledge, no surveys utilizing data from company respondents on the current state of practice in AI ethics exist. Existing surveys have relied on document data, for example from guidelines or project documents. Such surveys have been conducted on tools and methods [14], AI ethics guidelines [12] , Artificial General Intelligence projects [3]. Various such document-based surveys also exist on the technical side of AI development, such as on machine and deep learning techniques and tools. Respondent data have been utilized in surveys on public opinions [8], as well as surveys on evaluating AI ethics guidelines [16], but not the state of practice AI ethics specifically.

To provide needed insight into the current state of practice in the industry, we present survey data from 211 software companies. Our data provides some context for this special issue by helping us understand where we currently are as an industry in terms of AI ethics. For practitioners, the data can also serve as a way to benchmark where your organization stands.


- *V. Vakkuri is with the University of Jyväskylä, 40014, Jyväskylä, Finland. Email: ville.vakkuri@jyu.fi.*
- *KK. Kemell is with the University of Jyväskylä, 40014, Jyväskylä, Finland. E-mail: kai-kristian.o.kemell@jyu.fi.*
- *J. Kultanen is with the University of J yväskylä, 40014, Jyväskylä, Finland. E-mail: joni.kultanen@jyu.fi*
- *P. Abrahamsson is with the University of Jyväskylä, 40014, Jyväskylä, Finland. E-mail: pekka.abrahamsson@jyu.fi.*


xxxx-xxxx/0x/$x.00 © 200x IEEE Published by the IEEE Computer Society

## 2 What is AI Ethics?

Much of the research on AI ethics up until now has been predominantly theoretical and conceptual; valuable work aiming to define what is AI ethics (e.g. [5]). This has mostly been done by focusing on key principles [11]. These principles focus on specific categories of practical issues related to AI ethics, such as accountability. In our survey here, we focused on transparency, accountability, and responsibility, as well as predictability as a subset of transparency. Except for predictability, these three principles comprise the so-called ART principles for AI ethics.

Transparency is about understanding how the system works [7]. This is both about transparency of algorithms and data, the technical side of the system, but also transparency related to the development of the system [1], [10]. Transparency in terms of data and algorithms is related to the idea of explainable AI systems. Aside from being able to understand the system, we should also be able to understand who made the system into what it is today, and why.

Predictability can be considered a subset of transparency. It is about having a system that does what we expect it to do [2]. We certainly expect our autonomous thermostat in a smart home to keep the room temperatures in comfortable levels despite what it learns about our habits.

Accountability refers to liability issues related to stakeholders: who is liable to whom, for what, and why? To this end, laws and regulations can also be considered to fall under accountability. [1], [7], [10].

Responsibility is vaguer. It is about acting ethically or doing what we feel is the right thing. It is not tied to any specific idea of morality. [7]

Finally, Fairness, though not touched upon in our survey, is about equality in AI systems. Fairness has been discussed in terms of fairness in data or bias, as well as in terms of who benefits from AI systems [9], [10]. For example, do AI systems widen the societal gap between technologically skilled individuals and those less skilled?

Though these principles have focused on recently, various others have also been discussed. For example, the recently published European Union (EU) Ethics Guidelines for Trustworthy AI [1] considered trustworthiness to be the goal AI systems should aim for. The guidelines treat trustworthiness as a higher-level principle that principles such as transparency are required to achieve. Other principles include, for example, data-related ones such as privacy [11].

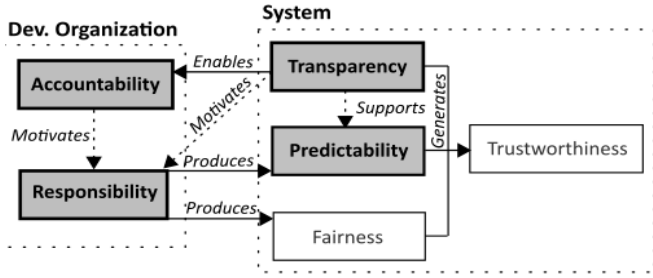


Fig. 1. Relations of key principles in AI ethics [19]

Fig. 1 portrays the relations between the principles we focus on. We have focused on the highlighted themes in our survey. Moreover, transparency is only considered in terms of data and algorithms in the figure.

Bringing this discussion and these principles into practice has been an ongoing challenge in the area [4]. For the most part, attempts at bridging this gap have been made by producing guidelines for AI ethics. The most prominent ones have been IEEE's Ethically Aligned Design (EAD) [10] guidelines. Other notable AI ethical guidelines include the European Union (EU) Ethics Guidelines for Trustworthy AI [1]. Overall, various guidelines have been produced by larger industry players, standardization organizations, academia, and governments alike [11]. These guidelines have various other principles than the ones we have chosen to focus on as well, such as data privacy, non-maleficence, and human well-being [11].

We currently have no knowledge of what impact these guidelines have had in the industry, however. Similarly, the current state of practice of AI ethics in general remains unknown, which is something we now shed some light on in this article.

## 3 And What is Actually Happening in the Industry?

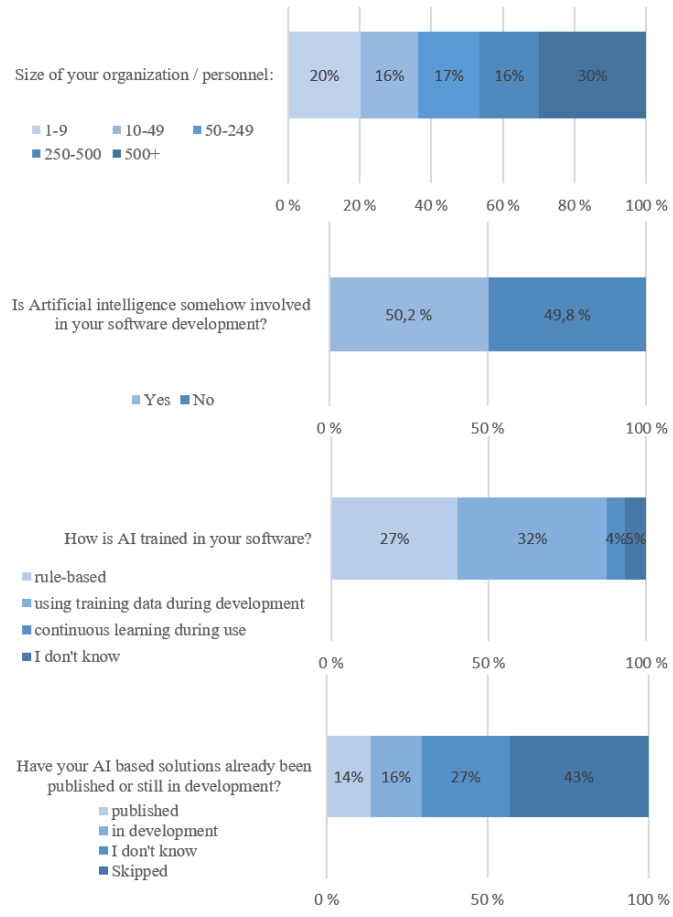


Fig. 2. Demographic description of the companies

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MS.2020.2985621, IEEE Software



Has the public and academic AI ethics discussion had an impact? Have these guidelines been adopted by the industry? To provide insights into the current state of practice in AI ethics, we conducted a survey, gathering responses from 211 software companies. The respondents were largely individuals capable of influencing the development in their companies: 68% of the respondents answered 4 to 7 in response to the question "how much can you personally affect the functionalities of the software developed in your organization and decisions made on them?".

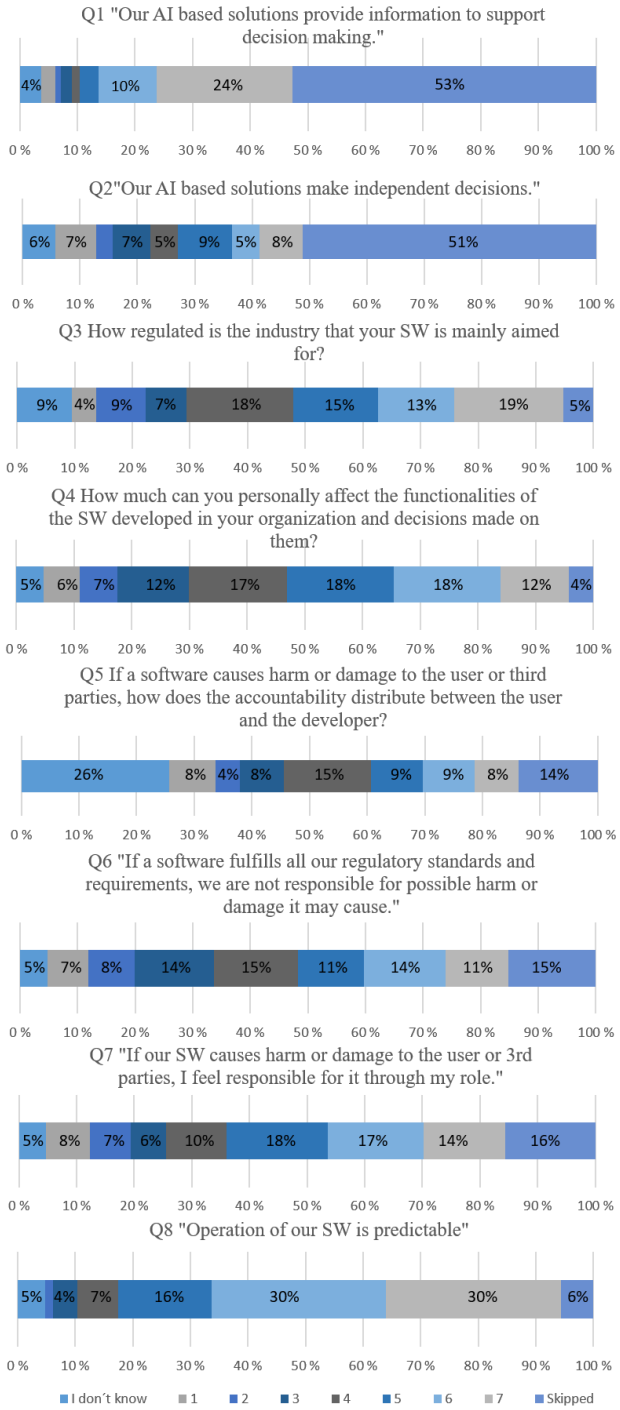


Fig. 3 The developer, liability, and responsibility. Scale from Strongly disagree to Strongly agree. Q3 Scale from Not at all to Very regulated. Q4 Scale from Not at all to Full authority

A little more than half of these companies (Fig. 2) were either developing or deploying an AI system. However, the responses did not notably differ between the companies that did not develop AI and the ones that did. We therefore included all responses. This is an interesting observation in and of itself: AI is currently simply treated as a feature in terms of ethics. This is in line with a study that argues that 90% of the activities we do in AI projects are the same as in any software project [17].

Overall, the responses indicated mixed maturity in implementing AI ethics. Responses to some of the questions directly indicated immaturity in relation to AI ethics, while some indicated some maturity. It would appear that the various AI ethics guidelines have not had a notable impact on practice, as has been suspected to be the case [4].

As many as 39% of the respondents skipped or answered "I don't know" to the liability question (Q5)(Fig. 3). This points to this being an unfamiliar theme, and thus an overlooked issue from an ethical viewpoint. Moreover, the qualitative responses from the companies also indicated that they did not tackle these issues even as well as their responses to the likert scale questions would have made it seem otherwise.

On the other hand, in response to some questions, such as predictability, the companies indicated more concern towards AI ethics related issues. For example, half of the organizations (48%) had a fallback plan for irregularities. Many respondents nonetheless noted that they did not have a fallback plan in place for unexpected system behavior in place, or that they did not know whether they had one (Fig. 4). Interestingly, most organizations (51%), felt their system could not be misused.

The respondents felt that they could influence the development of the system(s) highly, but still outsourced responsibility to the users when asked whether the developer or user was responsible (Fig. 3). 36% of the ones that answered (Fig. 3) considered meeting mandatory regulatory standards sufficient in terms of responsibility; past that it was up to the (end-)user to stay safe. Aside from the responsibility of their company, 49% of the respondents (aside from the 16% who skipped the question) felt personally responsible for any harm caused by their software, even if they largely didn't know that who was ultimately the one responsible.

Meeting the mandatory regulatory standards was also considered sufficient in terms of documentation by 43% of the respondents that answered (Fig. 4). On the other hand, 26% simply reported documentation being scarce or there being no documentation at all. The idea of being able to trace decisions back to individuals which is often discussed with accountability was reportedly achieved by 43% of the companies. However, the qualitative answers of many these companies made us doubt whether they really did address accountability to this extent with their work practices.

Their responses to documentation also somewhat conflicted with how the companies considered transparency important (Fig. 4). Transparency seemed to not be considered in terms of transparency of systems development.

Moreover, transparency in terms of data and algorithms

was mostly considered from the point of view of the development team and to some extent from the point of view of the user. Few companies considered transparency to public authorities, with 19% simply answering "I don't know" to the question regarding it as well. Transparency to public authorities is one topic of discussion in AI ethics [6].

Despite machine learning being associated with an increased unpredictability, the responses between the AI companies and other software companies did not notably differ. By far the most respondents felt that their systems were predictable. Yet, 34% of the companies had also faced issues due to unexpected operations in the system, pointing to a possible contradiction.

As most of our respondents were either from Finnish or US companies, we also compared the data between these two locations. There were no notable geographic differences in the data. Primarily, the Finnish companies operated in more regulated industries, and consequently seemed to place more emphasis on adhering to industry regulations.

### The Survey

We collected survey data from 249 respondents in 211 software companies, out of which 106 developed AI systems. All responses were included together in the figures, as we noticed during the analysis that the trends were very similar whether the companies developed AI. Indeed, the original idea of the survey was to compare how much more well-versed in AI ethics AI companies were compared to other software companies. Given the increasing ubiquitousness of AI systems, every software company is likely to soon to be involved with AI.

The survey featured three types of questions: (1) demographic questions (organization size, name etc.); (2) Likert scale questions; and finally (3) open-ended questions. In this article, we focus on the Likert questions, which are covered in their entirety.

The survey focused on some of the central principles in AI ethics in the past few years. Namely, we discussed issues related to transparency, accountability, responsibility, and predictability. We have discussed the meaning of these principles in the second section of this article. Furthermore, we discuss the research model in detail in another research article [19].

In the Likert scale questions, we asked the participants to evaluate the importance of principles such as transparency. They were also posed some practical questions, such as whether they had faced issues with unpredictability in their software.

We collected data from both multi-national organizations as well as ones locally based ones. Most companies were either US (53) or Finnish (111). The rest were from 18 other countries. Responses were collected either as F2F structured interviews or via an online survey. US based company responses were obtained by purchasing the SurveyMonkey Audience service. Interviews were conducted when possible in terms of scheduling. Most of the responses were collected F2F.

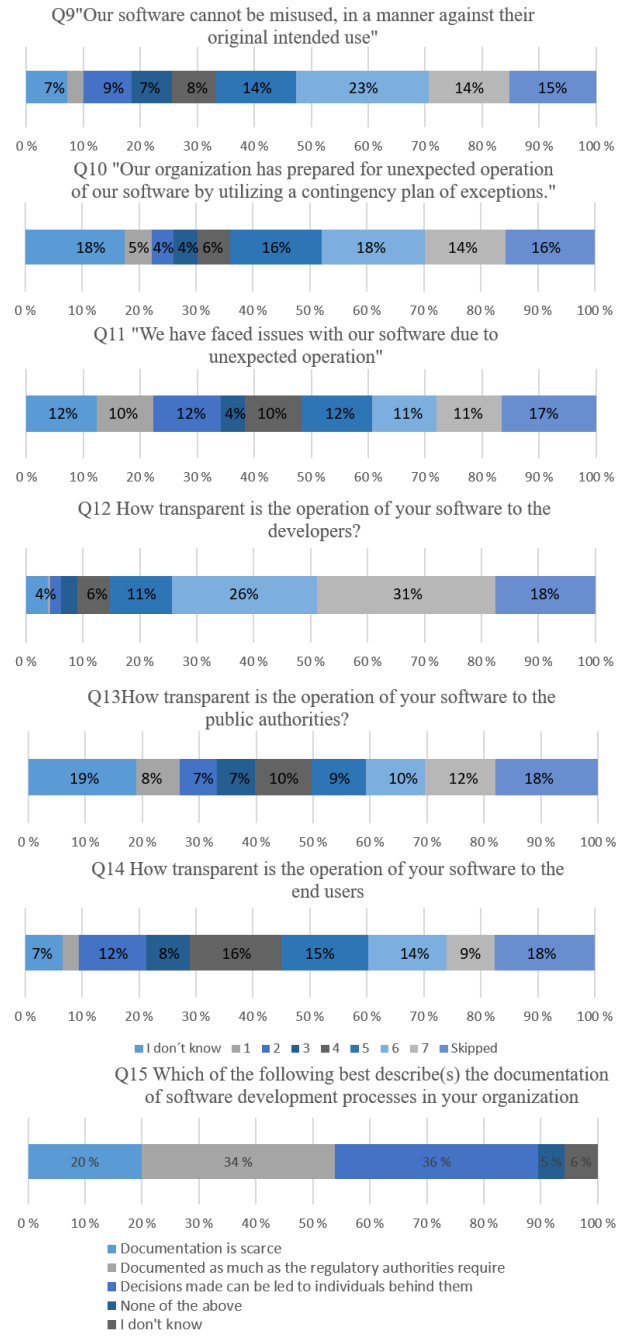


Fig. 4 Unexpected operation and transparency. Q9-Q11 Scale from Strongly disagree to Strongly agree. Q12-Q14 Scale from Not at all transparent to Fully transparent

## 4 What Should Your Organization Do?

The data we collected points to AI ethics implementation still being in its infancy. This observation is mostly based on how the companies developing AI had largely similar responses to the survey as the ones not developing AI. AI seems to be considered just another feature, at least as far as the ethical side of things is considered.

AI ethics is closely tied to other emerging ethical mega trends. Ecological issues such as data center electricity consumption are tied to the larger trend of being environmentally conscious. Similarly, data privacy issues are highly related to AI systems as AI systems typically handle vast amounts of data [1]. Regulations such as the General Data Protection Regulation (GDPR) are already forcing industry organizations to act in terms of data handling and have highlighted the interest of governments to tackle AI ethical issues. As your users become increasingly conscious about privacy issues, being ethical in relation to data privacy for example can become a selling point.

If you wish to implement AI ethics, guidelines such as the IEEE EAD [10] ones, among others [11], can provide a starting point. However, utilizing these guidelines requires additional work from your organization as they do not come in the form of an off-the-shelf method. You need to first make them more practical for your developers, project teams, and product owners and customers.

On the other hand, various tools for implementing AI ethics also exist [14]. However, unlike guidelines, which focus on the bigger questions in the design and development of the system, the currently available tools focus on small portions of the development process. For example, various tools to manage unpredictability in machine learning exist, but they only cover a small subset of AI ethics. Project-level methods for software development do not yet exist for AI ethics [14]. This is something research in the area is currently working to tackle [12]. As a startring point, we recommend focusing on certain key practices rather than relying solely on values and principles.

Ultimately, AI projects are, at least currently, like any other software project. According to a study [17], 90% of what is done in AI projects is the same as in any software project. AI development is still software development, and for that reason, developers play an important role in AI ethics as well. Product owners' responsibility is to make sure that sprint backlog items have ethical user stories included. From the software development viewpoint, ethics in AI could be viewed as a non-functional requirement of an AI-based software system. When it becomes tangible, it becomes more manageable.

Finally, in implementing ethics in AI, there are some antipatterns to avoid:

- Outsourcing ethics, for example to a high-level ethics committee. Quality in software development cannot be outsourced and neither can ethics.
- Assuming ethics can be successfully implemented without doing so systematically. Leaving ethical issues for the developers to tackle is unlikely to work. With no methods to help them, developers are left to rely on their own capabilities.
- Appointing one individual to implement ethics. No one person can or should do it. AI ethics is a strategic matter. For example, the whole development team should be involved, going back to what we mentioned in the previous paragraph.

Currently, few laws and regulations that force the industry to implement AI ethics exist. However, with regulations such as the GDPR being drafted globally, preparing to tackle AI ethics issues already is insurance for the future. Much like how adding pipes to an already finished house is far more expensive than adding them while it is being built, ethical issues are much cheaper to tackle during design or even development than deployment.

Even without being forced to do so, devoting resources towards tackling ethical issues such as transparency can already be beneficial for your organization. When you increase the level of documentation in the name of transparency, you also support stakeholder communication. In this fashion, AI ethics can produce benefits. From the point of view of AI ethics, stakeholder communication is important particularly in relation to the general public and regulatory authorities. You can also learn valuable lessons from past incidents such as the two mentioned in the introduction.

As AI systems continue to become even more widespread, the number of such incidents, large and small, will only grow. The software industry is in a key position in preventing this from happening. Acting on AI ethics today will quickly pay back.

## References

[1] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics guidelines for trustworthy AI," https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. 2019.

[2] M. Ananny and K. Crawford, "Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability," New Media & Society, vol. 20, no. 3, pp. 973–989. 2018.

[3] S. Baum, "A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy (November 12, 2017)". Global Catastrophic Risk Institute Working Paper 17-1. http://dx.doi.org/10.2139/ssrn.3070741. 2017.

[4] M. Brent, "Principles Alone Cannot Guarantee Ethical AI". Nature Machine Intelligence. 2019.

[5] J. Bryson and A. Winfield, "Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems," Computer, vol. 50, no. 5, pp. 116-119. 2017.

[6] V. Charisi, L. Dennis, M. Fisher, R. Lieck, A. Matthias, M. Slavkovik, J. Loh, A. F. T. Winfield and R. Yampolskiy, "Towards Moral Autonomous Systems," Preprint arXiv:1703.04741. 2017.

[7] V. Dignum, "Responsible Autonomy," Preprint arXiv:1706.02513. 2017.

[8] European Commission, "Autonomous Systems - Report". Special Eurobarometer 427 / Wave EB82.4 – TNS Opinion & Social, https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_427_en.pdf. 2015.

[9] A. W. Flores, K. Bechtel and C. T. Lowenkamp, "False positives, false negatives, and false analyses: a rejoinder to Machine bias: there's software used across the country to predict future criminals, and it's biased against blacks," Federal Probation, vol. 80,

no. 2, 38-46. 2016.

[10] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition," IEEE. Available at <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html>. 2019.

[11] A. Jobin, M. Ienca and E. Vayena (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, no. 1, pp. 389-399. 2019.

[12] J. Leikas, R. Koivisto and N. Gotcheva, "Ethical framework for designing autonomous intelligent systems". Journal of Open Innovation: Technology, Market, and Complexity, vol. 5, no. 1. 2019.

[13] L. E. Lwakatare, A. Raj, J. Bosch, H. H. Olsson and I. Crnkovic, "A Taxonomy of Software Engineering Challenges for Machine Learning Systems: An Empirical Investigation," In Proceedings of the International Conference on Agile Software Development, pp. 227-243. Springer, Cham. 2019, May.

[14] J. Morley, L. Floridi, L. Kinsey and A. Elhalal, "From what to how. an overview of AI ethics tools, methods and research to translate principles into practices," Preprint arXiv:1905.06876. 2019.

[15] Reuters, "Amazon scraps secret AI recruiting tool that showed bias against women," https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G. 2017.

[16] L. Rothenberger, B. Fabian and E. Arunov, "Relevance of Ethical Guidelines for Artificial Intelligence - A Survey and Evaluation". In Proceedings of the 2019 European Conference on Information Systems (ECIS). 2019.

[17] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Philips, D. Ebner, V. Chaudhary, M. Young, J. F. Crespo and D. Dennison, "Hidden technical debt in machine learning systems," Advances in Neural Information Processing Systems. 2015.

[18] The New York Times, "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far". https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html. 2018.

[19] V. Vakkuri, K-K. Kemell, J. Kultanen, M. Siponen and P. Abrahamsson, "Ethically Aligned Design of Autonomous Systems: Industry viewpoint and an empirical study," Preprint arXiv:1906.07946. 2019

**Ville Vakkuri** is a PhD student in Information Systems at the University of Jyväskylä. Contact him at ville.vakkuri@jyu.fi.

**Kai-Kristian Kemell** is a PhD student in Information Systems at the University of Jyväskylä. Contact him at kai-kristian.o.kemell@jyu.fi.

**Joni Kultanen** is a MSc student in Information Systems Science at the University of Jyväskylä. Contact him at joni.m.kultanen@jyu.fi.

**Pekka Abrahamsson** is a full professor of Information Systems and Software Engineering in the Faculty of Information Technology at the University of Jyväskylä. Contact him at pekka.abrahamsson@jyu.fi.