

This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Please cite the original version: Nevanperä, M. ; Rajamäki, J. & Helin, J. (2021) Design Science Research and Designing Ethical Guidelines for the SHAPES AI Developers. Procedia Computer Science, Volume 192, 2330-2339.

doi:10.1016/j.procs.2021.08.223

Available at: <https://doi.org/10.1016/j.procs.2021.08.223>

[CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)



25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Design Science Research and Designing Ethical Guidelines for the SHAPES AI Developers

Minna Nevanperä, Jyri Rajamäki* Jaakko Helin

Laurea University of Applied Sciences, Vanha maantie 9, 02650 Espoo, Finland

Abstract

This article targets the design process of ethical guidelines for the SHAPES project (Smart and Healthy Aging through People Engaging in Supportive Systems) which is a H2020 Innovation Action project. The aim of the project is to build solutions that can make it easier for older individuals to live at home, such as, robots, wearables and sensor technologies that apply artificial intelligence (AI). The guiding method of the design process of ethical guidelines is Alan Hevner's Design Science Research. Theoretical background consists of a form of literature overview, which contains the most relevant ethical theories and research on AI ethics, machine ethics and human rights. This article introduces the process of building the ethical guidelines for the SHAPES project and further discussion if providing guidelines is the sufficient tool to developers to take ethical action in development of the AI systems. The SHAPES guidelines include the following themes; accountability, transparency and explainability, diversity, inclusion and fairness, safety and security and societal wellbeing and humanity.

© 2021 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of KES International.

Keywords: Design Science Research; Ethical Guidelines; SHAPES; Ethical Competence

1. Introduction

The SHAPES project (Smart and Healthy Aging through People Engaging in Supportive Systems) is a European Union H2020 Innovation Action project. The aim of the project is to enable new types of operating models and markets through an open ecosystem. The project develops digital solutions for older individuals who are in some

* Jyri Rajamäki. +358 40 7642750

E-mail address: jyri.rajamaki@laurea.fi

way impaired or have illnesses that make their lives difficult. The aim is to build solutions that can make it easier for older individuals to live at home, such as, robots, wearables, sensor technologies. The purpose of the artificial intelligence-based ecosystem is to collect and analyze information on the needs of older people and to use this information to produce individual solutions to perceived aging-related problems. Technological or social analysis alone is not enough, but there is also a need to take into account the views of the target group, such as how artificial intelligence systems can have effect on good ageing and whether it can replace human care or reduce exclusion or loneliness. Perspectives can also be contradictory. What is effective and desirable for the society may not be desirable for the individual. As SHAPES project includes designing AI solutions therefore it is interesting to examine how to promote ethical competence of the developers and promote ethical behavior. It is also important to evaluate during the process if the guidelines are the best possible method to promote ethical system development. This article introduces the process of designing ethical guidelines according to Design Science Research and then demonstrates the results of this process and introduces the guidelines designed for the SHAPES project. Finally, there is some discussion on the significance of the ethical guidelines when promoting ethical awareness.

2. The Process of Designing Ethical Guidelines for the SHAPES project

This article applies Alan Hevner's Design Science Research (DSR) as its methodological background for constructing ethical guidelines for the SHAPES project. Designing ethical guidelines is considered as Hevner's theory's design artifact. The aim is to find useful methods for designing ethical guidelines for AI projects and to find the best way to give this kind of guidance to developers of the AI system. It is important to notice that DSR considers that the knowledge and understanding of the design problem and its solution are acquired when building the artifact. According to DSR, the outputs can be constructs, models, methods or instantiations [1]. When designing guidelines for a project, it is necessary to take into account not only the theoretical framework but also the operating environment, the people for whom the guidelines are designed and the specific features of the technology. In Design Science Research, all of these aspects become part of the design process.

Hevner has identified three Design Research Science cycles which are relevant when positioning design project into the wider context. The Relevance cycle brings in the context that the design research process. The cycle is also interested in bringing something back to environment during the designing process. Usually this is achieved by bringing in innovative artifacts that improve the environment. The Relevance Cycle not only define research process context, but also gives the criteria for acceptance of the research results and evaluation [2]. Rigor Cycle brings in the knowledge base and theoretical background into the designing process. Researcher needs to make sure that existing theoretical knowledge is taken into consideration in design process to ensure that the designs produced are research contributions, not only routine designs based on known design artifacts and processes [2]. The design cycle is to iterate between the design activities and evaluation of the artifact and the theoretical background and processes of the research. This can be seen as a heart of the design research process. Even though the design cycle draws from the other two cycles, it is important to understand that it is not dependent on other cycles [2].

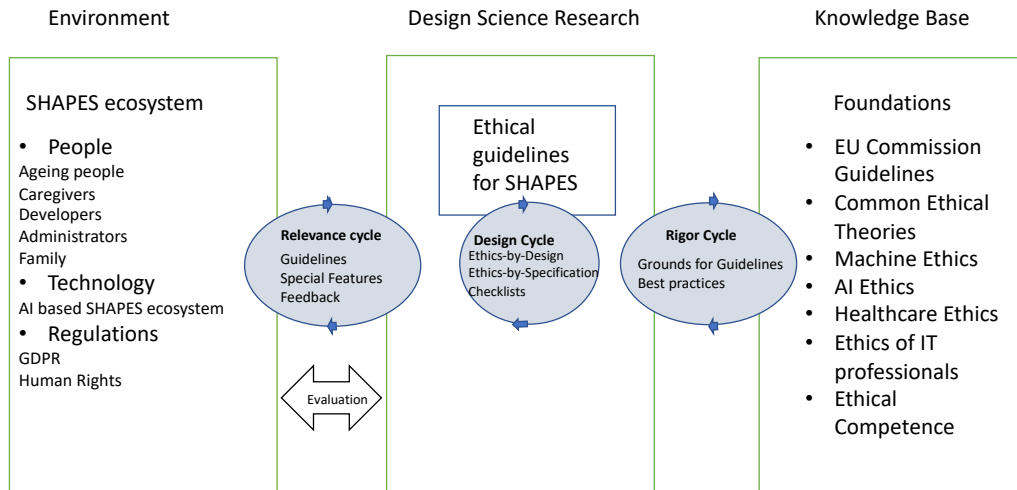


Figure 1. Design Science Research framework of the study. Modified from Hevner's DSR Framework [2].

2.1. Knowledge Base

The design process of ethical guidelines of SHAPES began with examination of the theoretical background such as common ethical theories, machine ethics and AI ethics. The knowledge base discussion also included ethical guidelines that already exist. The European Commission's Ethical Guidelines for AI [3] is one that might have a great relevance on the field since it is very thorough. To compare other ethical guidelines [4, 5, 6, 7] are studied to get a picture how the ethical issues are approached in commercial environment.

2.2. Environment

Design process of the guidelines included the examination of environment, in this case the SHAPES ecosystem. The special features of SHAPES that were examined were the ageing people, health care and the features of the artificial intelligence. The study also included examination of different approaches to how to design ethical IT systems. Ethics-by-Design and Values-in-Design were contemplated. These methods bring ethics into technical solution by giving the system ethical boundaries in which the system operates.

2.3. Designing Ethical Guidelines

Hevner's theory as a guideline many iteration rounds were performed between the guideline suggestion, the knowledge base and the environment. Because of the nature of the project Hevner's theory's Rigor Cycle became in this study more relevant than the Relevance Cycle. The Design Cycle evaluation was created more in relevance to theoretical consideration than actual field testing or having feedback from the system or people that are using the ethical guidelines.

As part of the information collection a few AI ethics seminars and lectures were attended. The most important knowledge was that the ethics of artificial intelligence is a very broad issue. First of all, concept of what is considered to be artificial intelligence is not always easy to define. Secondly, when considering the effects of the AI, it is fast changing technology that has new solutions and use cases appearing every day. The effects are broad and include technical matters as well as societal effects. The hype around AI is enormous and the expectations vary from AI to be the technology that saves the world to the technology that destroys the world. The potential is huge, but at the moment we are somewhere in between. When the field of study is broad also the ethical issues to be considered vary.

University of Helsinki introduced open education course on AI ethics in 2020. In this course they have valued five principles of ethics that are relevant in regarding to AI. These are, the principle of beneficence/ non-maleficence, accountability, transparency, fairness and respect of human rights. As we can see, the themes that are recurring in all discussion on AI ethics are fairness, transparency and accountability [8].

3. Ethical Guidelines for the SHAPES project

In this section the results of the design process for the SHAPES project guidelines are introduced.

3.1. Accountability

Even though machine ethics consider artificial intelligence at least partially as a moral agent, in the real life we cannot at least not yet hold artificial intelligence and its algorithms accountable for the decisions it makes. Especially in health care project like SHAPES the decisions that AI makes must be a subject of human evaluation. Accountability is strongly linked to the concept of responsibility. As discussed in machine ethics, philosophically moral responsibility requires that the moral agent is conscious of its own actions. This means that the agent should be able to evaluate and predict consequences of its actions. This means that AI at its current state cannot be held responsible of its own actions in moral sense. It is safe to say that human being must be accountable for all decisions that artificial intelligence makes. However, it is not always easy to set the criteria how the responsibility should be individualised. For example, we cannot name one person responsible on all effects of AI on society level. Although when we discuss on SHAPES project, we must be ready to at least project level to have a responsible person specified.

One of the European Commission's requirements is human oversight. In most guidelines this is linked to accountability. Human-in-the loop or human-on-the-loop approaches are mostly discussed, but also ethics-by-design can be seen as a resolution into the dilemma. According to Virginie Dignum et al. Ethics-by-design in AI is concerned with methods, algorithms and tools that are needed to ensure that autonomous agents with capability to reason take the path of ethical decisions and that their behavior stays within the moral boundaries that are given to the system [13]. Ethics-by design does not remove human oversight, but it gives boundaries to AI to make decisions without human oversight. When the decision is not within these given values, AI system cannot make the decision automatically on its own.

The guidelines for SHAPES project:

- 1) Name responsible person for all stages of AI development and use. Make it easy to contact when AI decisions are unexpected.
- 2) AI system should be developed with ethics-by-design-approach to provide ethical boundaries which within the AI system can make decisions.
- 3) Decide when the human-in-the-loop and human-on-the-loop-approaches are necessary. Keep in mind that information and decision-making in SHAPES includes health data.
- 4) Make it easy to overtake AI when it seems to generate unfamiliar decisions.

3.2. Transparency and Explainability

Transparency itself is not an ethical issue. It is just a matter of making operations and decision-making of the AI system visible. What can be seen as an ethical issue is that how much information the developers and operators of AI system are to give to the stakeholders and users. As users of the AI system, should we be aware that we are engaging with artificial intelligence and should we know exactly how the systems works? What is the sufficient information? Transparency links to human rights. A user has right to know on what grounds the decision is made and how the information is used in decision-making. For example, does the decision-making of the AI system violate privacy or right for self-determination if it tracks movements of the ageing person at home when the purpose of this is to prevent injuries? A user has to be aware what information AI system is using and how it is used in decision-making. If this information is not available, the decision-making of the system is not easy to be argued or disagreed.

We also need to discuss on which level of understanding of the algorithmic decision-making is sufficient. This is the matter of explainability, which is mentioned in almost all ethical guidelines. It might be because it is under discussion that what level of understanding of the AI system working logic is sufficient that we can use the AI system. Transparency can be divided into three components, simulatability (the understanding of how the model works), decomposability (understanding of the algorithmic components) and algorithmic transparency (visibility of algorithms) [8]. It has been discussed do the users need to have understanding or visibility on all three levels of the AI system or what level of understanding the developers and those who maintain the AI system must have? Can AI system be a “black box”, the system so complex that no one has full understanding on what basis the decisions are made? There has been discussion on making algorithms of the AI systems openly available. However, this will not solve the problem of comprehensiveness since the most users cannot understand the algorithms used in their coded form.

SHAPES-project needs to take into account the subject group, the ageing. The ageing people of the day do not always have the necessary IT-skills to understand algorithmic decision-making. This is why it is crucially important to make the AI system used transparent to all people, so that also the health care professionals or family of the ageing individual can view the decision-making process.

The guidelines for SHAPES project:

- 1) Design the AI processes reviewable and give the user the information that the decision is based upon.
- 2) Enable the user to access an explanation of why the AI system behaved as it did.
- 3) Design models for users and use visualization when communicating the algorithmic decision-making to stakeholders. Have the ageing people in mind.
- 4) Inform openly ways of using AI and be open for discussion of the working methods and development of AI.
- 5) Keep records of the development process and the decision-making.
- 6) Offer possibilities to users and other stockholders to give feedback.
- 7) Minimize the risks that might appear from transparency for example security risks.
- 8) AI's decision-making process should be explainable. The technology behind AI must be understandable for some, but how it comes to conclusions must be understandable to the most.
- 9) The user must have a choice whether he/she acts with a machine or a human being.
- 10) Make it easy to decide and interact.
- 11) Guide through the process.
- 12) The user must be always aware when and how he/she is interacting with AI.

3.3. Diversity, Inclusion and Fairness

Diversity, inclusion and fairness are ethical values that are not always easy to find solutions to. Regarding artificial intelligence, it can marginalise the people who are already in vulnerable position even more when they cannot access the resources that are needed to be able to benefit AI. On the other hand, artificial intelligence has great potential to promote well-being for these groups if they are able to access resources enabling the use of AI. In SHAPES the ageing is especially vulnerable group since not only they have difficulties and loss of abilities that ageing brings, but also that their ability to understand and use technology might be limited. If the ageing people have less competences to use digital technology and computers, the women in general are less likely to have these competences, the ageing women are at even greater risk to be excluded from technological evolution. One concern in AI development is also the gender divide. Only 12 per cent of the machine learning researchers are women. When the AI systems are developed and designed by men, there might be a risk for the gender-bias. Also, when men are more likely to adopt and use new technologies the solutions are more likely to be designed for them. When designing for older individuals, it is crucial that the development process includes the ageing from both genders [8].

One of the most important ethical issue to solve is fairness and discrimination. First of all, it is important to define discrimination. Usual definition for discrimination is that people belonging into certain group are treated differently from the people not belonging to this group [8]. It should be noted that this definition does not exclude possibility to positive discrimination. Another definition of discrimination is different treatment of the group by perceived membership that is causing social harm to this group [8]. This definition sees discrimination always harmful. Discussing discrimination brings us to discuss the concept of equality. Equality by definition means that everyone is treated the same way with no exceptions. But this usually means that the most vulnerable groups do not have the same possibilities than the groups with more power and resources. When discussing AI ethics, it is especially important to notice this. Equality might not always the best solution to promote equity. It is important to discuss how the fairness and equity is achieved without risking equality.

When discussing fairness, one of the issues arising is bias. Chavalarias and Ioannidis find in their study 235 biases that could affect research. The same kinds of biases can occur similarly when training algorithms if the data-sets used are imbalanced. Bias itself can be either positive or negative, but in the case of ethics more problematic is the case of negative bias since the positive bias might even be at least in some cases what is hoped for. Bias can be particularly problematic when it is not visible or when people are not aware of its existence. This is called implicit bias. For example, if the medical doctor has different views of racial groups, it might affect physician's decisions of treatment. Bias is especially troubling if it is affecting engineering and design. It is worth noticing that there is already action to counteract this bias in design. Universal design is a good attempt to develop things with that way that it is accessible for everyone [9].

Algorithms of AI are not immune to our society's biases. As they find patterns in the datasets and their learning is based on data, they are not free from our stereotypes and other discriminatory behavior when we use the data that our society produces. For example, facial recognition has been struggling to recognize non-Caucasian faces throughout its history. After ten years of developing AI systems, it still has trouble of recognizing Asian-origin faces.

How should we prevent AI systems to become prejudiced? The most effective way is to pay attention to the composition of the expert group that is developing and teaching AI. The more diverse the group is the more there are different views and recognition. The selection of datasets is crucially important. How we find a secure dataset that does not promote the same human prejudice and how do we even define the desired dataset and outcome?

There are a few measures that are attempts to tackle the problem of bias, but they all have also some disadvantages. For example, anti-classification means removing the data that is causing issues from the dataset for example gender or race. In some cases, this approach might cause more issues. For example, if we remove gender and race from the health data, we might lose the information that is relevant for the decision-making.

- 1) Keep in mind the norms, values, experiences and gains of the user group and value the positive in them.
- 2) Be aware of the cultural differences, gender and age and be sensitive not to bias upon them. Use “design for all” if possible.
- 3) Review carefully the data used and use ongoing research and diverse data collection to minimize algorithmic bias.
- 4) If the bias is detected, investigate carefully to understand from where the bias is originated and how it can be removed.
- 5) Collect feedback from the users.
- 6) Create and use check lists which promote diversity.
- 7) Review and create possibilities to empower the groups that are more likely to be excluded from the resources.

3.4. Safety and Security

Safety and security are not ethical matters as such, but when considering them against human rights there might be some ethical discussion needed. When the goal is to protect individuals from the social, emotional and physical harm the safety and security are the norm. Requirement of safety in designing artificial intelligence systems is an obligation, not an option. The system must not only be safe and secure when it is working as expected, but it must be also safe and secure for the users when something unconventional happens.

Safety and security can be also be seen as technical safety and security. It is strongly linked into privacy and data protection, that are not discussed further here though. In SHAPES project also the user group, the ageing people, must be taken into account. They might have less experience of using IT-systems and have less ability to understand the security and safety measures needed. This means that the system should be created in that way that way that it needs as little as possible user activity to be safe and secure to use.

- 1) Safety and security must be taken into account in all phases of development and tested thoroughly before release. Provide documentation.
- 2) User group, the ageing, must be taken into account.
- 3) Promote easy access and conventions familiar to the ageing population.
- 4) Promote one user interface whenever possible.
- 5) Protect individuals from social, physical and emotional harm.
- 6) Create security measures also for the situations that unconventional action or malfunction happens.
- 7) Discuss openly the limits of the robustness, security and safety.
- 8) Set boundaries on which the AI system can work independently.
- 9) Discuss on what grounds AI can be used in health care. For example, must AI make more reliable decisions than human being or should we allow some false decisions?

3.5. Societal Wellbeing and Humanity

One of the essential parts of the SHAPES project is to find solutions that help ageing people to continue to live at home. At least in Northern Europe the loneliness of the elderly is an immense problem. For many older individuals the health care personnel might be the only human contact they have. When developing AI based solutions for their benefit, it is important that we do not replace the human contacts with technology. The aim should be that AI solutions created assist and support health care professionals so that their time with the ageing can be used more on human-to-human contact that on for example administrative tasks. AI solutions should also not only be developed for surveillance or gathering the health data on ageing people, the focus should be on wellbeing of the individual as a whole. Solutions should promote maintaining physical and social activity and promote ageing people to remain active also outside their homes.

- 1) Design and develop the AI systems in that way that human contact for the ageing will remain or increase.
- 2) Aim for diminishing loneliness. Promote social contacts between generations and peer groups.
- 3) Promote prevention in all forms.
- 4) Promote solutions that maintain functional capacity or increase activity
- 5) Development should aim to promote or maintain the relationships with nature and environment. Solutions should promote outdoor activities.
- 6) Development should aim for common good and benefit humanity

4. Adoption of AI Ethics in the Development Process

McNamara et al. reviewed how the ethical guidelines effect the work of the software engineers. They concluded that the ethical guidelines given had almost zero effect on the practices of the professionals. Unfortunately, the study did not examine the reasons why there was no effect. [10]. It has been also discussed that checklists like the one European Commission has provided, might instruct the development to focus only to the matters that are on the checklist, not the problems that might be completely new or that were not added on the checklist at hand.

4.1. Ethics-By-Design and Values-by-Design

It has been said that the technical infrastructures and technology often reveal human values mostly because of the tensions, failures and counterproductivity. Values-in-design (VID) approach as Nussbaum et al. it defines, tries to create discipline that includes values into socio-technical designing process from the beginning. According to VID approach the difference in definition of ethics and values is necessary. Ethics is seen as a set of prescriptions whereas values are seen as action. This is why the values-in-design approach is discussing about values, not ethics. Values-in-design approach values cooperation, co-creation and coordination as methods of creating more human friendly technology. It emphasizes the importance of stakeholder participation. [11]. There are also other disciplines that value the same approach designing technology. For example, Freeman and Nissenbaum call it value-sensitive-design, Acre calls it critical technical practice and Sengers et al. calls it reflective design. All of these have a same goal, bring values into development process from the beginning, but they also emphasize the role of the designer as a carrier of the values and the fact that the values of the designer affect the way that technology is imagined, how it handles data and what kind of decisions are made [12].

Katie Shilton studied ethnographically how the designers in CENS laboratory took ethical matters into consideration when designing sensor technology. Shilton noticed that the designers were often aware of the ethical issues of the system they developed might face, but the ethical matters were seen outside of the technical staff's expertise and as a job for someone else to do. In many cases the value-sensitive design process was seen unattractive since it was much slower. Values were not seen as important as functional system and values were often forgotten when they competed with the system offiency and practicality. However, when the designers prototyped the application, they were fully aware of the ethical issues they had, since they saw them in practice [12.]

Like earlier discussed Ethics-by-design in AI is concerned with methods, algorithms and tools that are needed to ensure that autonomous agents with capability to reason take the path of ethical decisions and that their behavior stays within the moral boundaries that are given to the system. This means that the AI system should behave in a way that it is beneficial to people and safe to use. AI ethics can be divided in two, regulation (legislation and standards) and design (system itself). Ethics-by-design is considering the latter. The most important question according to ethics-by-design approach is that are we able to and how to build AI systems that have ethically-aware agents? The key issue is to articulate that artificial intelligence requires researchers and designers to be able to translate human values and ethical considerations into technical requirements. Actually, the ethics-by-design approach is very near to values-in-design approach when discussing that designers must take mental shift towards thinking values before performance of the system. Ethics-by-design approach requires that the question about AI

system reasoning should be the priority over performance [13.]

According to Dignum et al. three ethical issues are particularly concerning to AI systems, accountability, responsibility and transparency. We can see that the European Commission is especially concerned with the same issues. Dignum states that these three values are important to discuss when trying to ensure the societal good for everyone. How the AI system is seen to follow these ethical principles, depends upon what kind of reasoning is seen to be possible for AI system. If we believe that the AI system is not capable of ethical reasoning, it means that we should always have human supervision. That also means that the supervisor should have sufficient knowledge and means to do this job. This approach is called human-in-the-loop. Another approach to ethical reasoning of the system is that the environment itself has been designed that way that the deviation is impossible and the moral decision making of the system is not needed. Ethics -by-design considers AI system as an ethical agent itself. These agents are known as artificial moral agents. That means that the AI system is able to include moral reasoning into its deliberation and decision-making and explain its behavior in terms of moral concepts. This approach requires complex decision-making algorithms based on deontic logics. The system design needs very explicit and complex design based on reinforcement learning to be able to act as a moral decision-maker [13.]

4.2. Ethical Competence

It was earlier discussed about the McNamara et al.'s study where they found no effect on practises when the ethical guidelines were introduced to software developers. In the light of this it was decided that it was not enough to this project to have only ethical guidelines, but more was needed. The one additional approach is the concept of ethical competence.

When considering the development of the SHAPES project and impacts it has, it can be noted that very different perspectives have to be taken into account when the impacts are considered. A mere technological or social analysis is not enough, but we also have to take into account the views of the target group, such as what is considered as good ageing and whether technology can replace human care or reduce exclusion or loneliness, for example. Perspectives can also be contradictory, what could be effective and desirable for society may not be desirable for the individual. Assessing ethical implications is not straightforward, as ethical values themselves are already diverse and different arguments lead to different factors being included in ethical values. In ethical problems, there is often not just one right answer, but often the solutions are incomplete. The study of the ethics of artificial intelligence is not well-established and has many features that influence ethical evaluation, even if they are not really matters of ethics. An example of this is information security.

The subject of how to promote ethical development of AI systems is interesting to study, as it can be assumed that technical developers already inherently have some kind of code of ethics, for example with regard to respect for human rights. It is important to measure whether the guidelines can guide developers to make better ethical decisions, or do the guidelines, for example, influence developers to take into account only the ethical aspects included in the guidelines and to ignore other emerging ethical issues? We earlier discussed Katie Shilton's ethnological research of technicians that were aware of the ethical issues, but thought that they are the job for someone else. Often ethical guidelines see AI ethics as a technical issue and usually the problem-solving is attempting to find a technical solution to ethical issues than trying to solve the economical or societal issues of the larger scale.

5. Conclusions

The ethical guidelines for the SHAPES project were designed by using Alan Hevner's Design Science Research Approach. Theoretical background and environment were leading the design process in which the ethical guidelines

were designed through the iteration process. Environment in this case included studying the specific requirements that the target group, the ageing, and the SHAPES ecosystem being a health care system brought into using artificial intelligence.

It can be argued whether the ethical guidelines are the best way to produce ethical awareness. It might be better to have some sort of combination of ethical guidelines, ethical training and promoting ethical competence by ethics rounds or some other methods that use real cases that arise from the work of the developers. Also, the technical methods like ethics- by-design or values-by-design are important when considering making artificial intelligence to act responsibly. One subject for the further studies could be to study the effectiveness of the ethical guidelines or to compare different methods of promoting ethical awareness.

The problem with ethical guidelines that are designed for the developers is the issue of the level of the ethical concerns. Many ethical issues related to artificial intelligence are not on the level of the individuals, but on the level of the society as a whole. Artificial intelligence has possibility to change our societies in ways that we cannot yet even imagine. This means that to solve ethical concerns related to AI should be discussed widely on societal level.

It is important to discuss also the possibility that AI might be able to make more accurate and less biased decisions without human intervention. Do we want AI to make decisions like human being even though we know that humans make bad decisions or do we want to AI make decisions like a machine can even though we do not always understand on what basis the AI has reached the decision? How we know it is a good decision? More discussion is needed when human intervention or guidance is needed and when there are robust enough technical solutions to let AI work independently without human intervention or decision-making.

References

- [1] Hevner, Alan, March, Salvatore T., Park, Jinsoo & Ram, Sudha (2004) “DESIGN SCIENCE IN INFORMATION SYSTEMS RESEARCH” *MIS Quarterly*, **28(1)** : 75-105. DOI 10.2307/25148625.
- [2] Hevner, Alan & Chatterjee, Samir (2010) “Design Research in Information Systems Theory and Practice” Springer New York. DOI 10.1007/978-1-4419-5653-8.
- [3] European Commission’s High-Level Expert Group on Artificial Intelligence: ETHICS GUIDELINES FOR TRUSTWORTHY AI. 2019. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- [4] Everyday Ethics for Artificial Intelligence, IBM design program office. 2019. <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf> Referred 27.4.2020
- [5] Artificial Intelligence at Google: Our Principles. <https://ai.google/principles/> . Referred 28.4.2020.
- [6] <https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimar6>. Referred 28.4.2020
- [7] IEEE: ETHICALLY ALIGNED DESIGN A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Referred 27.4.2020
- [8] University of Helsinki: <https://ethics-of-ai.mooc.fi/>. Referred 20.12.2020
- [9] Howard, Ayanna & Borenstein, Jason (2018) “The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity”. *Sci Eng Ethics*, **24**:1521–1536 <https://doi.org/10.1007/s11948-017-9975-2>.
- [10] McNamara, A., Smith, J., Murphy-Hill, E. (2018). Does ACM’s code of ethics change ethical decision making in software development? In G. T. Leavens, A. Garcia, C. S. Păsăreanu (Eds.) *Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering—ESEC/FSE* (pp. 1–7). New York: ACM Press.
- [11] Knobel, C. & Bowker, G. (2011) “Computing Ethics: Values in Design.” *Association for Computing Machinery. Communications of the ACM*, **54(7)**, p. 26. DOI 10.1145/1965724.1965735
- [12] Shilton, K. (2013) “Values Levers: Building Ethics into Design”. *Science, Technology, & Human Values* **38(3)**, pp. 374-397. DOI 10.1177/0162243912436985
- [13] Dignum, Virginia, Baldoni, Matteo, Baroglio, Cristina, Caon, Maurizio, Chatila, Raja, Dennis, Louise, Génova, Gonzalo, Kliess, Malte, Lopez-Sanchez, Maité, Micalizio, Roberto, Pavón, Juan, Slavkovik, Marija, Smakman, Matthijs, van Steenberg, Marlies, Tedeschi, Stefano, van der Torre, Leon, Villata, Serena, de Wildt, Tristan, Haim, Galit (2018) “Ethics by Design: necessity or curse?” (https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_68.pdf)