Designing the Trustworthy Principles of Artificial Intelligence - Case Finnish Police

Sanna Lassila

2021 Laurea

Laurea University of Applied Sciences

Designing the Ethical Principles of Artificial Intelligence - Case Finnish Police

Sanna Lassila

Leading Transformational Change

Thesis

November, 2021

Laurea University of Applied Sciences **Abstract**

Degree Programme in Leading Transformational Change

Master of Business Administration

Sanna Lassila

Designing the Trustworthy Principles of Artificial Intelligence - Case Finnish Police

Year 2021 Number of pages 66

The objective of this thesis was to produce ethical principles of Artificial Intelligence (AI) for the Finnish police. There is a constant pressure to digitalize public sector processes. Due to the increasing amount of digital information and requirements for efficiency, digital processes with AI aided solutions need to be taken in use.

Ethical principles of AI lead the way to the upcoming EU regulation on AI solutions. The principles are needed prior to the regulation to maintain the high level of trust that the Finnish police enjoys among the citizens.

The development task was to produce ethical principles that would support the needs of the police by taking into account the operating environment with applicable national laws and current regulations, while also considering the upcoming EU regulation.

Service design methods were used to design principles that would fit into the organizational processes and to the operative environment of the police. The methods included prototyping, in-depth interviews, and cooperative workshops. The development was done in phases. The first draft was sketched based on principles and requirements presented in the Ethics guidelines of trustworthy AI by European Commission. The first version evolved to second version by self-iteration, but the following versions got enhanced by interactions between different stakeholders.

The key findings of this thesis were that the service design methods used apply well in this kind of development of processes concerning a reasonably new topic. A tangible product, a prototype, serves as a basis for discussion and gives the interviewees or attendees of workshops a vision of the outcome, something to build on.

As the discussion of ethical use of AI is ongoing, the Finnish police wanted to act as a pioneer in law enforcement and publish their ethical principles of AI. The outcome of the work is a description of trustworthy principles of AI. The principles describe what is meant with AI in the police and how AI systems are used and what kind of measures implement the ethical application of artificial intelligence systems. Besides the ethical aspects, the trustworthy principles consider compliance with law and reflect the requirements of robustness.

AI systems are not ethical or unethical themselves, the issue concerns the use of the systems. Ethical principles guide the development, implementation and use of the systems. To apply the principles in practice, their requirements need to be communicated inside the organization and implemented to organizational processes. The implementation of such a small organizational change in practice is implemented and monitored in accordance with suitable change management processes.

Keywords: Artificial intelligence, AI, ethics, police, law enforcement, service design

Contents

1 Introduction

This thesis provides an overview on ethical aspects of artificial intelligence systems. The thesis first presents definitions and basic concepts of artificial intelligence and then continues with ethical aspects, considerations, guidelines, and the proposal for the European regulation on international and national levels affecting AI systems. Some current use cases where artificial intelligence is used in Finnish and European law enforcement are also presented.

After reflecting the background and knowledge base for ethical principles of artificial intelligence systems this thesis describes the research and design process of the ethical principles of artificial intelligence for the Finnish police.

1.1 Background

In the modern Finnish society, there is a constant pressure to digitalize and automate operations with the use of artificial intelligence. The amount of data has increased exponentially since 2010 and will keep on increasing (Holst, 2021). The Finnish police has been using artificial intelligence systems for a few years for example to recognize characters or compare facial images from a video recording to the police badge register. Many artificial intelligence systems used by public authorities affect fundamental human rights and because of the current lack of regulation some (international) use cases have been rather questionable.

This pressure for efficiency with the lack of regulation has raised a need to produce ethical principles for the development, deployment and use of AI systems in order to provide trustworthy AI systems. In Finland there are a few large companies and public organizations that have published their ethical principles of AI, but no organizations of the law enforcement were found to have published their ethical principles of AI in Finland nor worldwide.

1.2 Research problem

The Finnish police enjoys a high level of trust among the citizens. (Vuorensyrjä and Rauta, 2020). The aim and the research problem of this thesis is to produce ethical principles for the development, deployment and use of AI systems to support the trust that the Finnish police has. These ethical principles need to be suitable for the needs of the Finnish police for them to be applied in practice. The research problem of this thesis is to design the ethical principles for the use of the Finnish police to support the current and future needs of the police but also to support the upcoming European regulation on AI systems. The design of the ethical principles is conducted by using of methods typical to service design.

1.3 Objectives and scope

This thesis has two objectives. The first objective is to produce ethical principles for the development, deployment and use of artificial intelligence systems in the Finnish police. The principles should support the realization of the fundamental human rights, but also reflect on the processes and actual use of the of the current AI systems of the Finnish police organizations. They should also be written in everyday language.

The second objective is to provide a basis for the implementation of these ethical principles to support the upcoming EU regulation of AI systems. The requirements of these ethical principles will be taken as part of the current software development process to provide quality for development, deployment and use of artificial intelligence systems. The implementation of these ethical principles into the software development process itself is not in the scope of this thesis.

2 Artificial Intelligence

Artificial Intelligence (AI) is a leap of development in a framework of digitalization consisting of a collection of different technologies, methods, and applications. Artificial Intelligence does not only affect technology, but it also affects the whole society and people. The technical aspects should therefore be examined from several other different aspects such as moral, ethical values, politics, justice, and economics. AI technology and AI systems can be used to benefit many areas and businesses like medicine, agriculture, traffic, teaching, commerce and markets, law enforcement and war. (Ailisto et al. 2018, 6.)

2.1 Definitions of Artificial Intelligence

A classic definition of artificial intelligence is known as “The Turing test”. Alan Turing (1912-1954), a British mathematician, proposed that if after a 30 minute conversation a person is unable to distinguish an artificial respondent from human it’s artificial intelligence. (Ginsberg 2012, 7.)

On the other hand, in his article "Minds, brains and programs" John Searle objects on the concept of artificial intelligence by arguing that solely communicating based on complicated rules without really understanding the meaning is not comparable to human thinking. He introduces an experiment of thought called “The Chinese room argument”; a person is sitting inside a room acting as a computer, and Chinese characters are fed into the room where the person compares them to a list with English counterparts and returns the text in English without knowing or understanding any Chinese at all. The message of the experiment is that Turing’s tests communication is possible to achieve based on syntax but without semantics

(meaning of the symbolics) which only neurobiological processes are capable of. (Searle 1980.)

Stuart Russell and Peter Norvig consider AI through its possibilities as an intelligent agent and define AI "as the study of agents that receive percepts from the environment and perform actions", without trying to define the technical framework. (Russell & Norvig 2016, viii.)

Artificial Intelligence (AI) can also be defined by the aims of its use, whether the aim of the system is to think or to act. Most of the AI systems can be set under the following four categorizations:

- systems that think like humans
- systems that act like humans
- systems that think rationally
- systems that act rationally. (Russell and Norvig 2016, 2.)

The article 3 in the proposal for a Regulation on a European approach for Artificial Intelligence defines AI systems as systems developed with machine learning, logic- and knowledge-based or statistical approaches that "can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" (Proposal for a Regulation of the European Parliament and of the Council 2021, 39).

2.2 Data as a foundation for AI

The term of machine learning is often combined with artificial intelligence. Data plays a significant role with AI as machine learning systems are trained with data. Guided machine learning is used when the wanted result is known, unguided machine learning is the case when the data is searched for categories and reinforcement learning is about a non-human actor operating in a complex environment, for example self-driving vehicles. Algorithms then define how the data is processed and analyzed. Quality and ethicality of the data used for teaching the system is of cardinal importance. (Ollila 2019, chapter 3.)

2.3 Subfields of AI

Artificial Intelligence is a term composing of several different tools, technologies, and algorithms. Subfields of AI can be categorized by the different coverages of AI. Major areas for application of AI are Machine learning, Natural language processing, Expert systems, Machine vision, Robotics and Planning. (Kayid 2020, 1.)

Machine learning is a sector of information technology that uses statistical methods that enable computers the ability to learn from the data without specific programming. The AI

system is taught with a learning dataset guiding the algorithm to adjust its parameters and classify the data in order to create forecasts and make decisions. (Ailisto et al. 2018, 14.)

Natural language processing (NLP) means computer aided ways to produce and analyze natural speech. NLP includes subfields like translations of text and speech, automatic recognition of speech, optical character recognition and intelligent text input. (Ailisto et al. 2018, 11.)

Expert systems – Donald A. Waterman (1986) described expert systems as follows "An expert system is a computer program that embodies expertise about a particular domain and can use symbolic reasoning techniques to solve problems in this domain; problems that would need the assistance of a human expert in the real world. An expert system should also be able to explain its conclusions". (Rijckaert, Debroey, and Bogaerts 1988, 487.)

Machine vision systems have the functions of image acquisition, image analysis and/or the recognition of features or objects within an image as well as exploitation and imposition of environmental constraints. (Labudzki, Legutko, and Raos 2014, 903.)

Robotics is the physical dimension of AI. Robots are machines that are guided by computer software and are able to influence physically to their environment by their seizer, arms, rollers or feet. Robots can have sensors and be able to sense through their machine vision or sense of touch. (Ailisto et al. 2018, 19.)

Planning - The goal for an AI planning system is to generate a plan to solve a specified problem. The plan will be provided to a system defined for solving the problem, for example a robot. (Hendler, Tate, and Drummond 1990.)

2.4 Opportunities and risks of AI

Artificial Intelligence is developing fast and there is a need to take advantage of it. It will change the way we live by improving healthcare, increasing the efficiency of farming and different production systems through predictive maintenance, contributing to climate change mitigation and adaptation, increasing the security, and in many other ways. At the same time, with AI comes along a number of potential risks, such as opaque decision-making, different kinds of discriminations, intrusion in our privacies or risks of our data being used for criminal purposes. This is why ethical considerations and regulations are needed in the development, deployment and use of AI systems. (White Paper On Artificial Intelligence - A European approach to excellence and trust 2020.)

Artificial systems are already widely used in our societies, the questions to focus now is how, where, when and by whom AI systems are used. AI systems can be used to provide opportunities by fostering human nature and augmenting its potentialities but can also be

misused or overused causing risks or even underused (creating costs) out of fear of misusing or overusing. (Floridi et al. 2018, 690-691.)

AI systems can be used to support human flourishing by enabling the self-realization. Routine tasks can be handled by AI offering more time to humans to be consumed more relevant and sensible tasks. The risk in this kind of use is the deskilling of humans in skill-sensitive domains like healthcare or equalization of people when social structures and jobs are changing. Another way to use AI systems is to use them to enhance human agency. The whole society can get advantages on people working better and more efficiently aided by AI systems. Therefore, it is essential to focus on the responsibility of what kind of AI systems are used and how they are used and what are the advantages and benefits shared. The risk is the lack of such responsibility. AI systems can also help to increase societal capabilities of people or societies at large. AI can radically help reinventing societies to perform more efficiently or in a more sustainable way. The risk here is the lack of human control, important tasks or decisions should not be fully delegated to AI systems. The human control and the ability to monitor the performance of such AI systems should be fostered. Lastly AI systems with their data-intensive, algorithmic-driven solutions can provide opportunities in cultivating the societal cohesion by supporting collaboration and designing societal frameworks in issues with coordination complexity, like climate change. The risk is that by designing these societal frameworks AI systems may erode human self-determination or lead to unwelcome changes in human behavior. Altogether, the use of AI systems can provide our societies significant opportunities. Considerations on anticipating, avoiding and minimizing the risks will improve the possibilities to provide AI systems that promote human dignity and flourishing. (Floridi et al. 2018, 691-694.)

2.5 Role of AI in law enforcement

According to the study requested by the LIBE committee (Committee on Civil Liberties, Justice and Home Affairs) of European Parliament in the recent years the role of AI systems has increased in the field of law enforcement. AI systems have been seen to benefit and be of substantial use in law enforcement functions in the European Union. (Fuster and Brussel, 2020, 10).

The use of automated processing techniques and algorithms in crime prevention is growing while massive data sets are being processed faster and more accurately. (Algorithms and human rights - Study on the human rights dimensions of automated data processing techniques and possible regulatory implications 2018, 10).

The renewed personal data act of the Finnish police (616/2019) entered into force on 1 June 2019. The law allows the data obtained by the police under duty to be used for other purposes concerning investigation, detection, or prevention of crimes. It is the purpose of the

data that specifies the possibilities of use instead of the register and its ownership. This renewed act supports digitalization and creates possibilities for a broader use of data in a comprehensive database for the use of AI systems. (FINLEX ® - Säädökset alkuperäisinä: Laki henkilötietojen käsittelystä… 616/2019, 2019).

In the subchapters below are descriptions of the four main themes of AI in the European law enforcement according to the study "Artificial Intelligence and Law Enforcement, Impact on Fundamental Rights". The main themes of AI in the European law enforcement are all based on machine-learning. The themes are predictive policing, facial recognition, AI and criminal justice and AI and borders. (Fuster and Brussel 2020, 21.)

2.5.1 Predictive policing

Predictive policing is mainly based on algorithmic processing and analyzing of (big-)data sets. The methods of predictive policing are also used in prevention of crimes as investigation methods. Predictive policing methods are considered effective and productive since processing and analyzing of large data sets with AI consumes a lot less of resources than if the work was done by humans. Methods of predictive policing can be used to forecast the places and times of increased risk of crimes or even to predict future offenders or victims of crimes. In the prevention and detecting of money laundering and terrorist financing, predictive policing methods of artificial intelligence are widely used in Europe. They are used in obligations imposed on banks. In the field of law enforcement, AI systems involved with money laundering and terrorist financing have more to do prevention instead of prediction. (Fuster and Brussel 2020, 22-23.)

In 2021 the Finnish police is implementing an AI solution to investigate money laundering (Vuoden 2020 rahoitushaku 2020). Register plates have been controlled from police vehicles in the traffic for several years now (Rekisterikilven lukulaitteesta tuli kolmessa vuodessa poliisin paras kaveri – "Kaikki kehuvat", 2017).

2.5.2 Facial recognition

Facial recognition is about comparing two or more digital images and identifying people in the images with the use of AI technology. Facial recognition is one of the most disputed areas of AI. Images of people are regarded as biometric identifications and according to the EU data protection law are considered to deserve special protection. (Fuster and Brussel 2020, 24.)

According to the Act on the Processing of Personal Data by the Police, the police can use automatic facial comparisons to prevent, detect or investigate crimes. Character information in police records originally collected for crime prevention can be used to compare facial images. For example, images of wanted persons can be compared to camera recordings held

by the police. Automatic face image comparison can only be used within the limits of police powers. (Usein kysytyt kysymykset poliisin henkilötietojen käsittelyä koskevasta lainsäädännöstä, no date)

It is notable that according to the renewed Act on the Processing of Personal Data by the Police (616/2019), the use of passport and id-card images is prohibited for the means of crime prevention. This prohibition of use originates from the fact that the original purpose of the passport and id-card images does not derive from crime prevention nor detection or investigation. Only photos originating from registries of criminal investigation purposes can be used in recognizing on criminals. This limitation of the use is likely to have been made to protect citizens from possible error hits. (Usein kysytyt kysymykset poliisin henkilötietojen käsittelyä koskevasta lainsäädännöstä, no date).

The European Data Protection Boards guidelines 3/2019 define the use of facial recognition systems as use of biometric data and should therefore be respect to the principles of lawfulness, necessity, proportionality and data minimization as set forth in the GDPR ('Guidelines 3/2019 on processing of personal data through video devices', 2020).

Most of the debates and disputes about the ethicality and lawfulness of using facial recognition systems concern the "live-time" use of image feeds as it can have major implications for the realization of fundamental human rights (e.g. the efforts of China to identify members of Uighur minority). In UK the High Court of England and Wales declared the comparing of live-time feeds of facial images to a predetermined watchlist of people lawful as a means to locate the people in the watchlist. (Fuster and Brussel 2020, 24-25.)

The coalition of Public Voice, a coalition established in 1996 to promote public participation in decisions concerning the future of the internet, has editorialized on the use of facial recognition systems. The Public Voice published a declaration called "A Moratorium on Facial Recognition Technology for Mass Surveillance" in October 2019. The declaration "calls for a moratorium on the use of facial recognition technology that enables mass surveillance" appealing to the observation that facial recognition enables the development of semi-autonomous processes minimizing the roles of humans. (Ban Facial Recognition – thepublicvoice.org, 2021.)

Another question raising interest with the facial recognition systems is the use of private company technologies, like Clearview AI Facial Recognition System, that matches individuals to images from multiple different databases in multiple public platforms in the internet (Some Observations on the Clearview AI Facial Recognition System – From Someone Who Has Actually Used It …, 2020).

In April 2021 the National Police Board of Finland notified the Data Protection Authority about a facial recognition program, Clearview AI, piloted in 2019 by the National Bureau of Investigation. The National Bureau of Investigation was unsure to where the images entered into the facial recognition program had ended up. (Tietosuoja | Krp kokeili kasvojentunnistussovelluksen käyttämistä, testauksesta tehty ilmoitus tietosuojavaltuutetulle 2021.)

Since May 2020, the Finnish Police has used a facial recognition system to identify and capture criminals. The system searches for suspects by comparing certain pictures from the registry of police to for example video material from surveillance videos. According to Yle news police inspector Pekka Sallinen states that the system has proved to be useful in investigation. (Poliisi on saanut rikollisia kiinni kasvoja tunnistavan tekoälyn avulla ja haluaisi laajentaa valtuuksiaan – testasimme, miten kone toimii 2020.)

2.5.3 AI and criminal justice

The theme of AI and criminal justice includes the use of AI in the implementation of justice in due process but also the AI systems to facilitate the access to law (e.g. use of chatbots) (Fuster and Brussel 2020, 28).

The use of these kinds of predictive systems is critically dependent on the availability of data and according to the in-depth study accompanied in the CEPEJ Charter, it seems that predictive algorithmic tools are not widely in use in European courts (Ronsin and Lampos 2018, 14). Artificial intelligence technology could provide many possible use cases for the purposes of criminal justice. Regulatory framework establishing minimum standards for transparency and intelligibility, as well as governance framework considering potential effects of the incorporation of an algorithm into a decision-making process, and relevant ethical issues, is needed (Babuta, Oswald and Rinik 2018, vii).

In the late 2020 the Finnish Police got funds for the implementation of a chatbot to serve customers in non-urgent counseling in the poliisi.fi webpages. The implementation project of the chatbot should be completed by the end of 2022. (Vuoden 2020 rahoitushaku 2020.)

2.5.4 AI and borders

Many EU systems operating with algorithmic processing are typically connected to law enforcement authorities, such as Europol and national authorities like police, boarder control and customs. The aim of the co-operation between EU countries is to increase of both centralised and decentralised information systems. (Fuster and Brussel 2020, 29.)

In the EU there are a number of initiatives concerning the use and development of IT systems with algorithmic processing of personal information related to travelling, migration and

asylum purposes (Fuster and Brussel 2020, 29). According to the study and final report on the impact on fundamental rights profiling functionalities some concerns were raised in both existing and upcoming EU information systems ('Fundamental rights review of EU data collection instruments and programmes - FINAL REPORT', 2019). Below are two examples of profiling systems with some issues concerning ethical processing of data.

PNR stands for Passenger Name Records and the EU PNR system holds the information of air traffic passengers travelling within or out of the EU (Statewatch, 2019). The EU wide PNR system was set up by a directive n 2016. Originally the PNR data is collected mostly by third party private companies and transmitted to national authorities from where it ends up in the European PNR system. The discussion about the substantive requirements for the processing of PNR data in terms of fundamental rights, for example by assessing the PNR data and profiling passengers, resulted in July 2017 when the Court of Justice of the European Union stated that "the pre-established models and criteria should be specific and reliable making it possible to arrive at results targeting individuals who might be under a 'reasonable suspicion' of participation in terrorist offences or serious transnational crime and should be non-discriminatory". (Fuster and Brussel 2020, 31.)

ETIAS is a system under development and stands for European Travel Information and Authorization System and it concerns travel authorizations for non-EU residents travelling to Schengen area. The data filled in by the applicants is compared by an algorithm called 'ETIAS screening rules' to the data in the ETIAS Central System and to other databases and watchlists and to some specific risk indicators pointing to security, illegal immigration or high epidemic risks. Ethical aspects and the implementation of fundamental rights are ensured in the use of ETIAS system by ETIAS Screening Board, which is an advisory function and consults for its part Central Unit of Frontex and the ETIAS Fundamental Rights Guidance Board. (Fuster and Brussel 2020, 32-35.)

3 Ethical norms of the Finnish police and ethical guidelines of Artificial Intelligence

The following subchapters include the ethical codes of the Finnish police and different guidelines and considerations of ethics in the development and use of AI systems. The sources included in the chapter are either from cited researchers or International and European publications that have pioneered in developing of ethical guidelines.

3.1 The ethical codes of the Finnish police

The Finnish police enjoys high trust among citizens. Ethical codes of conduct of the police derive from the values of the police and are part of police education. At the end of police training, an ethical oath is sworn. The ethical codes of conduct consist of eight principles.

These ethical codes of conduct guide the work of the police and strengthen good governance and public confidence in the Finnish police. The ethical codes direct reflection on rights and responsibilities as a member of the work community. The ethical codes embody the values of the police, which turn into actions in practice. (Eettinen koodisto, no date)

1. **The operation of police is predictable, and its premises are always based in the law.** Citizens must be able to trust the police in all situations. For this reason, the actions of the police are predictable and must be based on the law. The police never resorts to illegal means in the performance of its official duties. The police ensures the realization of fundamental and human rights in all its activities.
2. **The police operates equally.** The work community of the police does not tolerate discrimination or harassment on any grounds. The police treats every person, both customers and colleagues, equally, regardless of their personal characteristics or background.
3. **The police is open and pro-public.** The police always act as transparently as possible, thus implementing the principle of openness. On the other hand, the police makes sure that it never leaks confidential information to the public. The information systems are used only to perform official duties and a high level of information security is take care of in all situations.
4. **The behavior of the police is exemplary**. The police are subject to a special duty of conduct by law, which is observed in official and leisure time. The police does not behave in a way that would compromise the confidence in its operations.
5. **The police is independent and autonomous.** The police do not pursue personal interests in its official duties and disability provisions are taken into account in all activities. The police is non-corrupt and does not accept any financial or other benefit that could jeopardize the confidence in the police. The most important thing is that citizens and society trust the police.
6. **Police personnel work in a community.** The police value good work community skills. A sustainable basis for well-being at work is created through the cooperation of all personnel. The police know how to adapt operations to change and are constantly ready to embrace something new. The most important resource is a work community that works together and shares know-how.
7. **Police values are reflected in leadership.** Leading of the work is done fairly, responsibly, and ethically. Ethical principles are not compromised in quick decisions and actions. The management includes monitoring the legality and quality of operations. The experience, skills and views of individuals are valued, regardless of their role.
8. **Police uses its resources appropriately, efficiently, and responsibly.** The human resources at our disposal, as well as the funds, assets and other resources allocated to

> our operations, are used in the most appropriate and efficient manner possible, in accordance with the needs required. The use of funds is transparent and documented.
>
> (Eettinen koodisto, no date)

3.2 Ethics in Artificial Intelligence Systems

Virginia Dignum, professor at the Department of Computing Science at Umeå University, argues in her article "Ethics in artificial intelligence: introduction to the special issue" that artificial intelligence systems have developed fast from being perceived only as tools to being perceived as autonomous agents. This development has raised many questions and considerations on the role and responsibilities of AI systems. In order to gain trust and be able to use AI systems efficiently in our society ethical consideration is needed to produce guidelines and regulations concerning the use of AI systems. (Dignum 2018, 1.)

The need for ethical consideration concerning the use of AI systems includes questions like: In which kind of situations AI can make the decisions? Who is responsible for the decisions made by AI? Will there be moral, societal, or legal consequences from the use of AI? How should AI systems be designed, and should they be controlled or should there be regulations concerning the design, development and use of AI systems? (Dignum 2018, 1.)

These needs for ethical considerations and questions of development, responsibilities and use of AI systems have led to the development of several initiatives concerning ethical AI. These initiatives have raised conversation and created guidelines for the ethical development and use of AI. (Dignum 2018, 1.) Among others these initiatives include Universal Guidelines for Artificial Intelligence from The Public Voice (The Public Voice, 2018) and Ethically Aligned Design by IEEE initiative on Ethics of Autonomous Systems (IEEE, 2019) as well the Ethics Guidelines for trustworthy AI by European Commission. (*Ethics guidelines for trustworthy AI*, 2019).

In her article about ethics in artificial intelligence Virginia Dignum states "*Theories, methods, algorithms are needed to integrate societal, legal and moral values into technological developments in AI, at all stages of development (analysis, design, construction, deployment and evaluation).*" She also argues that "Responsible Artificial Intelligence is about human responsibility for the development of intelligent systems along fundamental human principles and values, to ensure human flourishing and wellbeing in a sustainable world" and that "responsibility is fundamental to autonomy and should be one of the core stances underlying AI research". Accordingly, she argues that artificial intelligence is related to three levels of ethics:

1. **Ethics by Design** –Technical aspects of ethical AI systems. For example: How is it guaranteed for a human to have the possibility to interfere to the decision-making process of an AI system?
2. **Ethics in Design** – Regulations and engineering methods guiding the development of AI systems. For example: How it is guaranteed that AI systems making decisions are transparent, fair and do not discriminate?
3. **Ethics for Design** – The considerations and implementations, such as standards and codes of conduct, to guideline the ethical development and use of AI systems. (Dignum 2018, 2.)

According to Maija-Riitta Ollila (2019, chapter 3), philosopher and Doctor of Political Science, describes in her book "Tekoälyn etiikka" autonomy and adaptivity as two pivotal characteristics of AI ethics. Autonomy as for the ability to perform tasks independently in a complicated environment and adaptivity as for the ability to increase the performance by learning from the experience. Their importance is underlined because artificial intelligence systems are relatively new applications of technology and are still largely unregulated and in many respects our own ethical intuitions are lacking.

3.2.1 Ethical Principles of Artificial Intelligence Systems

AI4People, an initiative aimed to draft a set of ethical guidelines for AI, also did research in the field of applied ethics analyzing several statements and documents containing ethical principles. As a result, the research offered "a synthesis of existing sets of principles produced by various reputable, multi-stakeholder organizations and initiatives". The study found that the four principles of bioethics (**beneficence**, **non-maleficence, autonomy** and **justice**) adapt well with digital ethics and ethics of AI, needing to add only one more principle: **explicability**. (Floridi et al. 2018, 695-700.)

The principle of beneficence consists of the terms of "well-being" and "common good". AI systems should be used to promote the well-being of people and the planet. The principle of non-maleficence means that while promoting the objectives of well-being and common good AI systems should also prevent any intentional or unintentional harms arising from the use of AI systems. Such harms could be infringements on personal privacy. The principle of autonomy is a bit more complex when considering AI systems as many of the system are developed to work on their own; a balance should be found between the human decision-making power and the power delegated to artificial agents. The principle of justice is about promoting shared benefits and shared prosperity while preserving of solidarity between people by eliminating discrimination. The principle of explicability creates accountability on the processes of AI systems by enabling the understanding of the decision-making processes. (Floridi et al. 2018, 695-700.)

In 2019 a study called “Artificial Intelligence: the global landscape of ethics guidelines” listed 84 documents from private and public organizations around the world containing ethical principles or guidelines on artificial intelligence. The research team analyzed the 84 documents and as a result found that AI ethics merge to a set of main five principles: “transparency, justice and fairness, non-maleficence, responsibility/accountability, and privacy” (Jobin, Ienca, and Vayena 2019, 3-7). These five principles are considered to be the leading principles of AI ethics.

The principles of “justice and fairness” and “non-maleficence” align with the previous explanations of ethical principles by AI4People above. According to this study of Jobin, Ienca and Vayena the principle of transparency is often seen to be connected to the principle of explainability and is about minimizing harm by increasing the disclosure of information, interpretability or other acts of communication. The principles of responsibility and accountability can be considered as principles underlining acting with integrity and clarifying the attribution of responsibility and legal liability. The principle of privacy is seen “as a value to uphold and as a right to be protected”. It is also connected to data protection and data security. (Jobin, Ienca, and Vayena 2019, 7-10).

3.2.2 Human rights and Ethically Aligned Design (EAD)

In his research article Mark Latonero states “In order for AI to benefit the common good, at the very least its design and deployment should avoid harms to fundamental human values. International human rights provide a robust and global formulation of those values.” This pinpoints to the importance of the design and deployment of the AI systems. The implementation of the fundamental human rights should be taken into account already in the design phase of the system and continue in the implementation and use. (Latonero 2018, 5-6.)

The Universal declaration of human rights (UDHR) describes a set of 30 fundamental human rights promoting the dignity, equality, freedom, security and respect of all people (United Nations, 1948). These fundamental human rights are the foundation for the Ethically Aligned Design (EAD) promoted by a global initiative called the Institute of Electrical and Electronics Engineers (IEEE). (IEEE 2019.)

IEEE works to provide pragmatic and directional insights of ethically aligned design to serve the process of development, education, and policymaking of AI systems. The aim of the IEEE initiative is to promote the development of ethical AI systems by human-centric structures to benefit human society and the environment without infringing fundamental human rights or dignity and by increasing of human flourishing and environmental sustainability. AI systems should be developed and used to promote “humanitarian and sustainable development resulting in increased individual societal and environmental well-being”. (IEEE 2019, 2-4.)

The global initiative of IEEE has published a broad seminal work “Ethically aligned design” (EAD) to support the implementation of ethical AI. The treatise describes ethically aligned design framework to consist of three main pillars reflecting the anthropological, political, and technical aspects:

1. Universal human rights
2. Political Self-Determination and Data Agency
3. Technical Dependability. (IEEE 2019, 10.)

The treatise of EAD also presents eight general principles that define “imperatives for the ethical design, development, deployment, adoption, and decommissioning” of AI systems. These eight principles provide help in the deeper analysis and pragmatic implementation of AI systems. The principles are presented above as they are described in the IEEE treatise of Ethically Aligned Design:

1. **“Human rights** – “A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.
2. **Well-being** - A/IS creators shall adopt increased human well-being as a primary success criterion for development.
3. **Data agency** - A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people’s capacity to have control over their identity.
4. **Effectiveness** - A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.
5. **Transparency** - The basis of a particular A/IS decision should always be discoverable.
6. **Accountability** - A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.
7. **Awareness of Misuse** - A/IS creators shall guard against all potential misuses and risks of A/IS in operation.
8. **Competence** - A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.” (IEEE 2019, 11.)

The eight general principles map with the three main pillars as described in the Figure 1. The general principles implement the values of the main pillars of EAD. An ethical AI system covers all the main pillars by implementing the general principles. (IEEE 2019, 12.)

		EAD Pillars		
		Universal Human Rights	Political Self-Determination Data Agency	Technical Dependability
EAD General Principles	Human Rights	x	x	
	Well-being	x	x	
	Data Agency	x	x	x
	Effectiveness			x
	Transparency	x	x	x
	Accountability	x	x	x
	Awarness of Misuse			x
	Competence			x

Figure 1. EAD Pillars mapped to EAD General Principles. (Adapted from IEEE 2019, 12.)

Human rights are the foundation of ethical AI as can be seen from the IEEE treatise of EAD where Universal Human Rights are presented as the first of the main pillars and Human Rights as the first of the General Principles (IEEE 2019, 12).

3.3 The Berlin declaration on Digital Society

The Berlin Declaration on Digital Society is a document signed on the 8th of December 2020 by the EU Member State ministers. By signing the declaration, the member states engage in promoting the digital transformation and development of public sector digitalization with respect to the fundamental human rights and shared European values. The Berlin Declaration commits the member states to the effective, human-centered and sustainable use of digitalization promoting European values and rights. (European Commission, 2020).

The member states engage to lead by example with the digital development of their public authorities and adopt to cornerstone principles in the digital sphere. By adopting the cornerstone principles, they commit to work on these concrete objectives implementing the European values:

1. "Promote fundamental rights and democratic values in the digital sphere;
2. Enhance social participation and inclusion;
3. Foster digital empowerment and digital literacy;
4. Strengthen trust through security in the digital sphere;
5. Strengthen Europe's digital sovereignty and interoperability;
6. Create value-based, human-centered AI systems for use in the public sector;
7. Foster resilience and sustainability"

(Berlin Declaration on Digital Society and Value-Based Digital Government 2020.)

3.4 Realization of AI ethics in practice

As the initiatives and organizations are forming guidelines, principles, and codes of conducts for the ethical development and use of AI systems, there are findings that despite the good intentions, ethicality is not realizing in practice. The findings show that in order to make the ethical codes of conduct for AI systems to actually guide the development and use of AI, they must be put into practice so that they are present in all development, implementation, deployment and utilization of AI systems. (McNamara, Smith, and Murphy-Hill 2018, 732.)

A multiple case study by Vakkuri, Kemell & Abrahamsson researched how a tool called 'RESOLVEDD strategy' affected on the ethical consideration of the development team. RESOLVEDD strategy is a step-by-step decision-making tool intended for teaching practical ethics and first introduced by Pfeiffer and Forsberg (1993). The key finding of the study was that simply the presence of an ethical tool increased the ethical considerations in the development. (Vakkuri, Kemell, and Abrahamsson 2019.)

An increasing amount of AI systems are entering the markets. The pressure to process data more efficiently, also in industries and operations dealing with people's fundamental rights and private data, combined with the current lack of regulation, the interest and need for ethical guidelines is visible. An initiative called AlgorithmWatch, a non-profit Berlin-based organization, is committed to watch and analyze the automated decision-making systems and their impact on society. The AlgorithmWatch also maintains a repository of published AI ethics guidelines, with over 160 different guidelines collected so far. (AlgorithmWatch 2019.)

Another recent study from AlgorithmWatch called "The only way to hold Facebook, Google and others accountable: More access to platform data" indicates that transparency in the form of data access to journalists, academics and other civil society actors is needed to achieve an effective regulatory framework. Access to data and transparency of the data is crucial for the development of effective regulation and accountability mechanisms. (AlgorithmWatch 2020.)

3.5 International, European, and national guidelines and regulations concerning the ethical use of AI

In the beginning of 2021, there are yet no laws regulating the use and implementation of artificial intelligence systems in Finland. With the General Data Protection Regulation (GDPR), organizations are subject to requirements related to personal information systems, which must also be taken into account in the development of information systems (*Data protection under GDPR*, 2021). In the EU level there has been a growing need to prepare some regulations concerning the ethics on the use of artificial intelligence systems. Even though there are no binding regulations yet, there is a lot of prevailing work done. A proposal

by the European Commission for regulation concerning Artificial Intelligence was published on April the 21st, 2021. The aim of the proposal is to safeguard fundamental EU values, rights, and user safety by obliging high-risk AI systems to meet mandatory requirements related to their trustworthiness. The proposal will be a continuum to a white paper called "White Paper on Artificial Intelligence A European approach to excellence and trust". (European Commission 2021.)

While European Union has taken an emphasis on protecting the data and limiting the possibilities on exploiting the data, China's approach to the use of AI differs widely. China has implemented one of the broadest strategies for AI development "New Generation Artificial Intelligence Development Plan of China (2015– 2030)". The NGAI Development Plan draws a blueprint for a complete AI ecosystem for the country involving many functions of the society. "Governance Principles for the New Generation Artificial Intelligence — Developing Responsible Artificial Intelligence" is a publication issued within the framework of NGAI considering ethical development and use of AI systems. AI systems are already widely used in the everyday life of Chinese people connecting them to local businesses and financial services without any regulations affecting the use of AI systems. Chinese governments approach to the regulations is that it resolves first to evaluate the impact of AI technologies toward society by studying long-term and cross-disciplinary social experiments. Appropriate AI governance methods will be applied by these experiments. (Wu et al. 2020.)

The following chapters include some central guidelines that affect broadly the considerations on the ethical principles and lead the way for the upcoming legislation.

3.5.1 Public Voice guidelines

The Public Voice is a coalition formed in 1996 to promote public participation in decisions concerning the future of the Internet. The Public Voice has pursued issues concerning privacy and freedom of expression as well as the Internet governance. In October 2018 Public Voice published Universal Guidelines for Artificial Intelligence to inform and improve the design and use of AI and to maximize the benefits and minimize the risks while ensuring the protection of fundamental human rights. The Guidelines include twelve articles of which the primary responsibility resides with the ones funding, developing, and using the systems. The guidelines are listed below as they are presented in the Public Voice's Guidelines. (The Public Voice 2018.)

1. **"Right to Transparency.** All individuals have the right to know the basis of an AI decision that concerns them. This includes access to the factors, the logic, and techniques that produced the outcome.
2. **Right to Human Determination.** All individuals have the right to a final determination made by a person.

3. **Identification Obligation.** The institution responsible for an AI system must be made known to the public.
4. **Fairness Obligation.** Institutions must ensure that AI systems do not reflect unfair bias or make impermissible discriminatory decisions.
5. **Assessment and Accountability Obligation.** An AI system should be deployed only after an adequate evaluation of its purpose and objectives, its benefits, as well as its risks. Institutions must be responsible for decisions made by an AI system.
6. **Accuracy, Reliability, and Validity Obligations.** Institutions must ensure the accuracy, reliability, and validity of decisions.
7. **Data Quality Obligation.** Institutions must establish data provenance and assure quality and relevance for the data input into algorithms.
8. **Public Safety Obligation.** Institutions must assess the public safety risks that arise from the deployment of AI systems that direct or control physical devices and implement safety controls.
9. **Cybersecurity Obligation.** Institutions must secure AI systems against cybersecurity threats.
10. **Prohibition on Secret Profiling.** No institution shall establish or maintain a secret profiling system.
11. **Prohibition on Unitary Scoring.** No national government shall establish or maintain a general-purpose score on its citizens or residents.
12. **Termination Obligation.** An institution that has established an AI system has an affirmative obligation to terminate the system if human control of the system is no longer possible.“ (The Public Voice 2018.)

3.5.2 CEPEJ’s Ethical Charter on the use of AI

CEPEJ stands for European Commission for the Efficiency of Justice. In 2018 CEPEJ adopted the first European guidelines to ethical principles for the use of artificial intelligence in judicial systems. The guidelines are called “European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment”. The charter provides a framework of principles for the use of European policy makers, legislators, and justice professionals. The charter pays respect to the fundamental rights guaranteed in the European Convention on Human Rights (ECHR) and the Council of Europe Convention on the Protection of Personal Data. According to CEPEJ it is essential to ensure that AI remains a tool improving the efficiency and quality in the service of the general interest and that its use respects individual rights. (Council of Europe 2021.)

According to the charter, AI can “improve the predictability of the application of the law and consistency of court decisions, subject to compliance with the principles set out below. In criminal matters, their use must be considered with the greatest reservations in order to

prevent discrimination based on sensitive data, in conformity with the guarantees of a fair trial." (Council of Europe 2018, 5.)

1. **"Principle of respect for fundamental rights**: ensuring that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights.
2. **Principle of non-discrimination**: specifically preventing the development or intensification of any discrimination between individuals or groups of individuals.
3. **Principle of quality and security**: with regard to the processing of judicial decisions and data, use certified sources and intangible data with models elaborated in a multi-disciplinary manner, in a secure technological environment.
4. **Principle of transparency, impartiality, and fairness**: making data processing methods accessible and understandable, authorize external audits.
5. **Principle "under user control":** precluding a prescriptive approach and ensure that users are informed actors and in control of the choices made." (Council of Europe 2018, 7.)

3.5.3 European Commissions Ethics Guidelines for Trustworthy AI

In April 2019 European Commission published a document called The Ethics Guidelines for Trustworthy AI. The document was written by the High-Level Expert Group on AI (AI HLEG) and the aim of the document is to promote the development and use of Trustworthy AI as well as to provide guidance for AI applications in general in achieving Trustworthy AI. (Ethics guidelines for trustworthy AI 2019, 4.)

AI ethics is described as "a sub-field of applied ethics, focusing on the ethical issues raised by the development, deployment and use of AI" in the Ethics Guidelines for Trustworthy AI published by European Commission. The Ethics Guidelines for Trustworthy AI take a human-centric approach to AI ethics: the guidelines are reflected to and implement the Charter of Fundamental Human Rights in the European Union. (Ethics guidelines for trustworthy AI 2019, 9.)

The trustworthy AI consists of three, ideally in harmony working and overlapping components, which should be met throughout the system's entire life cycle:

1. **lawfulness**, complying with all applicable laws and regulations;
2. **ethicality**, ensuring adherence to ethical principles and values;
3. **robustness**, ensuring that the system performs in a safe, secure and reliable manner not causing any unintentional harm (technical and social perspective).

(Ethics guidelines for trustworthy AI 2019, 6-8.)

Fundamental human rights concerning AI systems

The Charter of Fundamental Rights in the European Union declares the universal values upon which the European Union is founded. It is a binding document that brings together the freedoms and rights of the European citizens. (Charter of Fundamental Rights 2020.) The Charter includes 54 articles reflecting to the modern European society considers aspects like personal data protection and freedom of expression and information (*Charter of Fundamental Rights of the European Union*, 2012).

The fundamental human rights were the basis for the creation of Guidelines for Trustworthy AI. Most of the fundamental rights are already legally obligatory but the ethical reflection can still help us to vision how AI implicates these fundamental rights and how it should be used instead of how it can be used. (Luotettavaa tekoälyä koskevat eettiset ohjeet 2019, 11-12.)

According to the Guidelines for Trustworthy AI the fundamental human rights particularly apt to cover AI systems are listed below:

1. **"Respect for human dignity** - AI systems should be developed in a manner that respects, serves, and protects humans' physical and mental integrity, personal and cultural sense of identity, and satisfaction of their essential needs.
2. **Freedom of the individual** - Human beings should remain free to make life decisions for themselves.
3. **Respect for democracy, justice and the rule of law** - AI systems should serve to maintain and foster democratic processes and respect the plurality of values and life choices of individuals. AI systems must not undermine democratic processes, human deliberation or democratic voting systems.
4. **Equality, non-discrimination and solidarity** - Equal respect for the moral worth and dignity of all human beings must be ensured with special attention on AI not producing unfairly biased outputs.
5. **Citizens' rights** – AI is to be used to improve the scale and efficiency of government in the provision of public goods and services to society at the same time making sure that citizen rights should be safeguarded" (Ethics guidelines for trustworthy AI 2019, 12-13).

Ethical principles of Trustworthy AI

Adhering to the fundamental human rights listed above, the AI HLEG developed four ethical principles, which strive to improve individual and collective wellbeing. The four principles derive from the fundamental human rights, to guide the development, deployment and use of AI systems. They are considered as ethical imperatives meaning that all AI systems and practitioners should strive at and adhere. Many of these are already regulated by law, yet the

adherence of ethical principles goes beyond formal compliance with existing laws. (Ethics guidelines for trustworthy AI 2019).

1. **The principle of respect for human autonomy** – AI systems should be designed to support human autonomy by augmenting, complementing, and empowering human cognitive, social and cultural skills. Humans must be able to keep self-determination over themselves and their decisions. AI systems should be human-centric and leave an opportunity for human choice.
2. **The principle of prevention of harm** – AI systems should not cause any mental or physical harm to human beings or to the natural environment.
3. **The principle of fairness** - The development, deployment and use of AI systems must be fair and free from unfair bias and discrimination and stigmatization. Equal opportunities to access to education, goods, and services should be fostered.
4. **The principle of explicability** – In order to gain trust among the users of AI systems, "the processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected". (Ethics guidelines for trustworthy AI 2019, 13-16.)

The guidelines take a human-centric approach to AI to foster the European values and principles deriving from the fundamental human rights. There may be tensions between the principles: For example, between *the principle of prevention of harm* and *the principle of fairness*, when AI is used to predict crimes in a way that collides with data privacy. Special attention should be focused to situations involving vulnerable groups such as children, persons with disabilities and other groups with risks of exclusion, and to situations which are characterized by asymmetries of power or information. (Ethics guidelines for trustworthy AI 2019, 16.)

Requirements of Trustworthy AI

The four principles described above are high-level ethical principles offering guidance. The principles are translated into concrete requirements, uniting the three components of trustworthy AI, lawfulness, ethics and robustness, to achieve trustworthy AI. The Ethics Guidelines for Trustworthy AI describe seven key requirements, listed here below, addressed to all stakeholders working with AI systems or using them. (Ethics guidelines for trustworthy AI 2019, 14.)

1. **Human agency and oversight** – The need for AI systems to support human autonomy and decision-making, as well as allow for human oversight and foster the fundamental human rights.
    - The realization of human rights should be verified by an assessment prior to the development of an AI system.

- Human agency should be ensured by allowing human interaction to a suitable degree and by ensuring that the end users do not remain subjects to a decision based solely on automated processing.
- Human oversight should always be ensured, and a human should always have the possibility to over-ride a decision made by an AI system.

2. **Robustness and safety** – The requirement of robustness and safety ensures cybersecurity. AI systems and software should be secure and reliable and thoroughly tested to mitigate risks of cyber-attacks.
3. **Privacy and data governance** – Privacy and personal data should be protected as well as quality, non-biased data should be ensured in AI systems. Citizens should have full control over their own data, and their data should not be used to harm or discriminate against them.
4. **Transparency** – The requirement of transparency is linked to the principle of explicability, and it ensures that AI is not biased. Traceability, explainability and communication are all elements of transparency. Traceability refers to documentation and that the AI systems should be traceable. Explainability refers to the need of AI systems to be understood and traced by humans. The element of communication means that humans need to be made aware that they are not communicating with a human.
5. **Diversity, nondiscrimination, and fairness** – The requirement links closely with the principle of fairness. A trustworthy and ethic AI system needs to enable inclusion and diversity as well as ensure equal access. To ensure the requirements of diversity, nondiscrimination, and fairness it must be ensured that the algorithms and the data are not biased. Stakeholders of the system that may be directly or indirectly affected should also be consulted and involved in the development and deployment of the system.
6. **Societal and environmental well-being** – The guidelines of trustworthy AI also call for social and environmental well-being grounding to the principles of fairness and prevention of harm. The environmental friendliness and social impacts of the system should be monitored and the effects of AI systems on society and democracy should be assessed.
7. **Accountability** – The requirement of accountability links to the principle of fairness and complements the other requirements. The responsibility and accountability of AI systems, especially concerning systems affecting fundamental human rights, and their outcomes should be assessed by internal and external independent audits. Reporting on the negative impacts of the AI system must be ensured and the reports should be made available. (Ethics guidelines for trustworthy AI 2019, 14.19.)

In case there are tensions between the requirements the trade-offs in fulfilling the requirements, decisions on the trade-off should be continuously evaluated. In such situations where there are no ethically acceptable trade-offs, the AI system should not proceed or be used in that form. To ensure trust accessible redress mechanisms should also be implemented. (Ethics guidelines for trustworthy AI 2019, 20.)

3.5.4 A European approach to excellence and trust

In February 2020 European Commission published a white paper concerning the development and use of artificial intelligence with a purpose setting out policy options on how to promote the uptake of AI and of addressing the risks associated with certain uses of this new technology. The white paper states “AI should work for people and be a force for good in society”. AI should be used to improve the wellbeing of European citizens and used to contribute solutions to some of the most pressing societal challenges linked to sustainability, demographic changes, protection of our democracies and to the fight against crime. The risks include “opaque decision-making, gender-based or other kinds of discrimination, intrusion in our private lives or being used for criminal purposes”. (White Paper On Artificial Intelligence - A European approach to excellence and trust 2020.)

According to the White Paper (2020, 1) “to address the opportunities and challenges of AI, the EU must act as one and define its own way, based on European values, to promote the development and deployment of AI”. It is the lack of trust that is the main factor holding back the development of AI systems. To fully seize the opportunities that AI offers some effort must be put to develop necessary industrial and technological capacities as well as to legislation and regulation to gain trust and guarantee the trustworthiness of AI systems. With the white paper the European commission launches a broad consultation to produce concrete proposals for a European approach to AI. (White Paper On Artificial Intelligence - A European approach to excellence and trust 2020, 9-10).

3.5.5 European Commission’s Proposal for Regulation on AI

The legislation process in the EU level consists of three institutions: The Commission presenting the interests of the Union as a whole and proposes new laws, and the Parliament representing EU citizens and the Council representing the governments adopt the new laws proposed by the Commission. The Commission together with the member countries then implement the laws. The Commission also ensures that the laws are properly applied and implemented. (*Institutions and bodies*, 2021).

There are different kinds of legal acts in the European union. Regulations are binding acts that are applied as such across EU member states. Directives are acts that set the goal to member states but leave the implementation of the laws to the countries. Decisions are

binding legal acts to those to whom they are addressed. (Regulations, Directives and other acts, 2016).

The European Commission's proposal for the regulation on artificial intelligence will be a binding act that regulates on the development and use of AI-based solutions while building trust and confidence towards AI systems at the same time. The ultimate aim of the proposed act is to increase human well-being with AI systems. (Proposal for a Regulation of the European Parliament and of the Council 2021.)

The vision for the European approach for AI regulation is that it should be robust and flexible at the same time, and not focus on technology as such. The legal framework should focus on concrete utilization of AI systems by not constraining nor hindering the technological development. (Proposal for a Regulation of the European Parliament and of the Council 2021.)

On the 21st of April 2021 European Commission published its proposal for the regulation on a European approach for Artificial Intelligence. The aim of the regulation is to provide trust to the use of artificial intelligence between the users and customers of public administrations and AI companies, it also aims at strengthening Europe's competitiveness and industrial basis in AI and unify national rules and regulations to prevent the fragmentation of EU's internal market. The proposed regulation concerns providers and users of AI systems within and outside of EU region when the systems affect persons located in the European Union. The objectives of the proposal are to:

- ensure that the AI systems in the European markets are safe and trustworthy respecting the laws on fundamental rights and European values.
- ensure legal certainty in order to support the innovation and investment on AI systems.
- enhance the governance and law enforcement on fundamental rights and safety requirements concerning AI systems.
- facilitate the development of trustworthy AI in the single market of EU and to prevent the fragmentation of the market. (Proposal for a Regulation of the European Parliament and of the Council 2021.)

The proposal for regulation takes a risk-based approach to AI systems. The proposal addresses risks concerned with AI applications, lists high-risk applications, sets requirements for high-risk applications, defines obligations for high-risk AI users and providers, proposes conformity assessments and enforcements for AI systems and governance structures at European and national levels. (Regulatory framework on AI | Shaping Europe's digital future, 2021).

The proposal divides AI systems in different levels considering the risks and harm possibly caused to the fundamental rights of natural persons. Risk levels of AI systems are based on

the purpose of their use, not based on the technology itself. Figure 2 describes the different levels of risks. Majority of the AI systems represent minimal or limited risk systems. (Regulatory framework on AI | Shaping Europe's digital future 2021.)

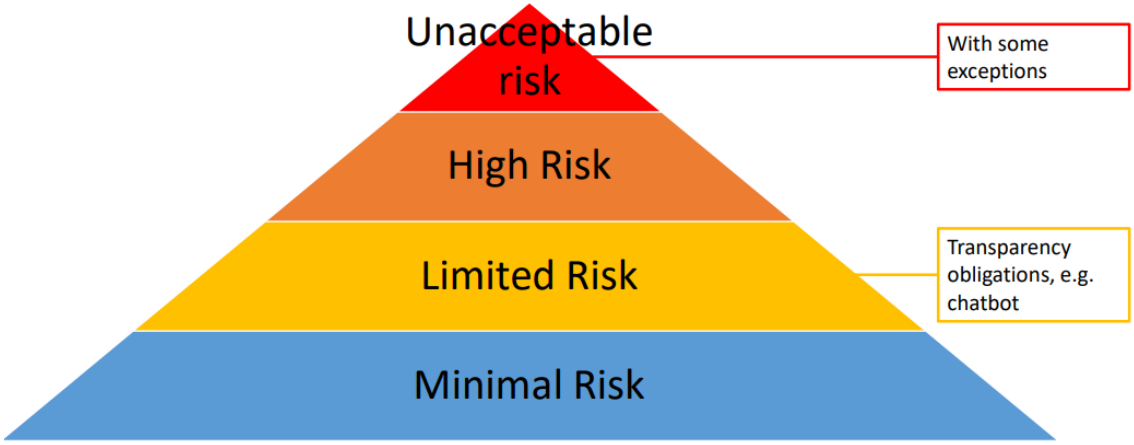


Figure 2. Risk based approach. (Adapted from Regulatory framework on AI | Shaping Europe's digital future 2021).

AI systems considered to cause unacceptable risks as clear threats to safety, livelihoods and rights of people will be banned according to the proposal. Examples of AI systems of unacceptable risks are systems manipulating human behavior, opinions or decision making, systems exploiting vulnerabilities of specific groups on users or user groups, systems evaluating or classifying the trustworthiness of natural persons providing social scoring of natural persons for the purposes of public authorities or systems performing real-time remote biometric identification in a publicly available place and spaces intended for the purposes of law enforcement. Such AI systems performing real-time remote biometric identification would only be allowed for the purposes of law enforcement with specific rules and regulations and objectives of "targeted search for specific potential victims of crime, including missing children", "prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack" and "the detection, localization, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA62 and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State". (Proposal for a Regulation of the European Parliament and of the Council 2021.)

Most of the regulations in the proposal regulate the development, implementation, deployment and use of high-risk AI systems described in the Annex III of the proposal. Briefly such high-risk systems are systems:

- performing “real-time” or “post” biometric identification and categorization of natural people;
- used as safety components in the management and operation of critical infrastructure;
- used to assess students or participants in tests required for admission to education or determine access to education or vocational training;
- used for recruitment, selection of natural persons for employment, making decisions for promotion and termination of contractual relationships concerning work or task allocation and for monitoring and evaluating the performance of employees;
- used by public authorities to evaluate the eligibility of people for essential services and benefits, or systems evaluating the creditworthiness of people (with some exceptions of small-scale providers in their inner use), or systems used to dispatch or establish priorities between people in first response services such medical aid.
- to be used for the purposes of law enforcement for making individual risk assessments, as polygraphs or to detect the emotional states of natural persons, to detect deep fakes unless authorized by law for the purposes of detection, prevention or prosecution of criminal offences, evaluation of the reliability of evidence, to predict occurrences of crimes based on profiling of natural persons, to profile natural persons in course of detection, investigation or prosecution of crimes or to analyze crimes by searching complex data sets from different sources identifying unknown patterns or hidden relationships in the data.
- to be used for the purposes of migration, asylum and border control management as polygraphs or to detect the emotional states of natural persons, to assess risks on natural persons entering EU, to verify travel documents or to examine applications or complaints for asylum, visa and residence permits.
- to assist in administration of justice and democratic processes. (Proposal for a Regulation of the European Parliament and of the Council 2021.)

The proposal regulates also on the quality requirements of the datasets used in the training of the high-risk AI systems. According to the proposal companies and organizations developing high-risk AI systems should have “access and use high quality datasets within their respective fields of activities”. The datasets shall be relevant, representative and of high quality without biases, assessed and subjects to data management processes. The proposal also empowers the Commission to update the list of high-risk AI systems. (Proposal for a Regulation of the European Parliament and of the Council 2021.)

In order to provide the requirements of accountability and transparency of trustworthy AI, the proposal also regulates on the testing procedures and requirements of the technical details and restrictions in the documentation, instructions of use, record-keeping and conformity

assessments procedures concerning high-risk AI systems. The outputs of the system need to be verified and traceable throughout the lifecycle of the system and the users of high-risk AI systems should be able to understand and control how system functions or makes the decisions. The human oversight of the high-risk AI systems is implemented by regulating on the tools or measures to enable human oversight and in an understandable way: the system must have built-in operational constraints to enable human intervention by stopping the operation or correcting the output. The requirements of robustness, accuracy and cybersecurity of high-risk AI systems are also regulated in the proposal. (Proposal for a Regulation of the European Parliament and of the Council 2021.)

Limited risk systems, like chatbots, have specific transparency obligations to make users aware of the interaction with a machine. Most of AI systems, like video games or spam filters, are considered as minimal risk systems with no obligations or regulations. (Regulatory framework on AI | Shaping Europe's digital future 2021.)

The proposal regulates as well on the obligations of providers, importers, distributors and users of high-risk AI systems with obligations of applicable assessments, documentation, instructions, monitoring and record-keeping (*Regulatory framework on AI | Shaping Europe's digital future*, 2021). Despite the fast-evolving nature of AI technologies, the proposal is designed to be future-proof by requiring ongoing quality and risk management procedures from the providers of high-risk AI systems even after the system is placed on the EU markets. (Proposal for a Regulation of the European Parliament and of the Council 2021.)

The proposal also proposes governance structures as new notifying authorities at European and national levels. The European Data Protection Supervisor shall as well provide AI regulatory sandboxes with supporting datasets to facilitate the development, testing and validation of new innovative AI systems for a limited time before entering the EU markets. Also an EU database containing information on stand-alone high-risk AI systems is to be set up. (Regulatory framework on AI | Shaping Europe's digital future 2021.)

The proposal announces an establishment of a new European authority "European Artificial Intelligence Board" contributing to the cooperation of national authorities and coordination on the emerging issues covered by the proposal. The requirements of conformity assessments and supervising duties of the use of high-risk AI systems in national levels requires either establishing of new notifying authorities or additions to existing authorities' obligations and responsibilities in order the proposal to enter into force. Yet (September 2021), there is not any definitive schedule for the enforcement of the proposed regulation, but according to the digital strategy of EU it could enter into force, by earliest, in the second half of 2022 with a transitional period of two years. (Regulatory framework on AI | Shaping Europe's digital future 2021.)

3.5.6 Finnish legislation and regulations concerning AI

The program of Prime Minister Marin's government includes an effort to promote digitization of the governance and society as a whole. The government program sets the goal of developing the regulatory environment and governance so that they enable digitalization and sustainable development, as well as a broad culture of experimentation. (Automaattisen päätöksenteon sääntelyä valmisteleva työryhmä asetettu 2021.)

In September 2021, there is no legislation available for regulating the use of artificial intelligence, but the renewed Act on the Processing of Personal Data by the Police (616/2019) regulates the use of personal data when processing of the data uses automatization. The preparation for national legislation on artificial intelligence has been started though. On the 18th of January 2021 Finland's Ministry of justice has set a working committee to prepare a general legislation concerning automatic decision-making of administration. The committee will be working with the preparation until 31.12.2021. The goal for the committee is to prepare the regulations at the general law level on the use of automation in administrative matters in making and preparing administrative decisions. Regulation should include the necessary regulations on the permitted use of automation, on information systems requirements, approval and control of information systems, transparency and public access to documents and other aspects of good administration, legal protection and other regulations necessary to ensure fundamental rights. (Automaattisen päätöksenteon sääntelyä valmisteleva työryhmä asetettu 2021.)

The renewed Act on the Processing of Personal Data by the Police (616/2019) entered into force on 1 June 2019. The new act applies the processing of personal data necessary for the performance of police duties when the processing is fully or partially automated or when the personal data constitute or are intended to constitute a personal register or part thereof. The laws are no longer tied to registers or information systems, but to the purposes for which personal data are processed.

The Act on the Processing of Personal Data by the Police (616/2019) gives the police the possibility to use the data obtained under their existing powers in a more versatile way in the prevention, detection, and investigation of crimes. The aim of the new Act (616/2019) is to ensure that the police processes personal data in compliance with the requirements of the new data protection legislation at the national and EU levels. (Poliisin ja Rajavartiolaitoksen henkilötietolait ja matkustajarekisteridirektiivin täytäntöönpanolaki voimaan 1.6. 2019.)

3.5.7 Summary of the guidelines and regulations

The guidelines and regulations discussed in the previous chapters are outputs of different levels affecting the Finnish police and the development of the ethical principles of AI. Figure

3 describes these different levels. The guidelines in the outer circle affecting to the development of ethical principles of AI for the Finnish police are universal or European non-binding guidelines. The middle circle describes the European regulations that are binding as such. The inner circle describes the national laws and acts that are binding and can apply over or complement EU regulations. Binding regulations and acts are written with bold font. Common to all of them is that they are designed to protect fundamental human rights.

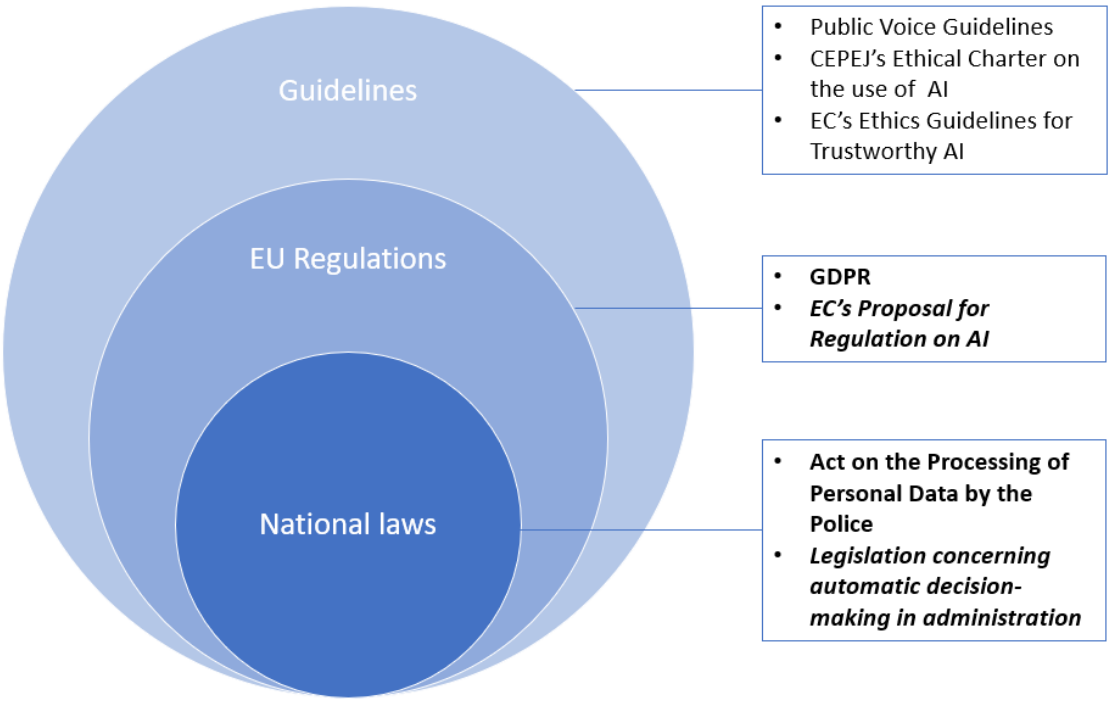


Figure 3. Levels of guidelines and regulations

4 Development of the ethical principles of artificial intelligence

The research problem of this thesis is to create ethical principles for the Finnish police to be used in the development of systems containing artificial intelligence. The principles should take into account laws and regulations, as well as best practices in accordance with general guidelines. The principles should also be written in everyday language to support the understandability and to increase the knowledge of the subject.

The idea for the thesis came from a concrete need of ethical principles and guidelines concerning the development of AI systems in the Finnish police. The Finnish police has its ethical codes presented above in the chapter 4.6. and these ethical principles and guidelines for AI should work in co-operation and complement the ethical codes of the police.

In 2020, the Finnish Police got special funding from the Ministry of Finance to support the process automation and digitalization using artificial intelligence. The funding is targeted to projects implementing a chatbot to support customer service, automated littering from speech to text and automated receiving and handling of money laundering notifications. The specification and development of these systems has begun and ethical guidelines are needed as soon as possible. (Vuoden 2020 rahoitushaku 2020.)

Developing and implementing of ethical principles of AI is an organizational change. Nadler & Tushman (1990, 80) categorize organizational changes to anticipatory or reactive and as incremental or strategic. This kind of a change in the organization's operations and processes that underpins the upcoming EU regulation is considered anticipatory and incremental as the change affect only selected components of the organization. To put into practice, these kind of anticipatory and incremental changes need just some tuning to the organizational practices instead of huge strategic recreations. (Nadler and Tushman 1990, 80-81.)

4.1 Research and development method

The research work is conducted by methods of qualitative research. Features of action research and grounded theory method using narratives were used as there was some previous research done on ethicality on AI, but not any publications on the creation of ethical principles for an organization.

The research work started in February 2021 by familiarizing with the literature about ethics and artificial intelligence and by completing a Massive Open Online Course (MOOC) "Ethics of AI" provided by the University of Helsinki. (Ethics of AI 2021.)

Through the literature and the MOOC "Ethics of AI" the basic concepts, dilemmas and declarations concerning the ethical use of AI got clarified and the study work concentrated on going through the different initiatives and organizations having published guidelines for the development and use of ethical AI or ethical principles of AI. The research problem conceptualized while going through the material: How to create ethical principles for AI systems that would truly be taken into practice and be part of the software development process and use of the systems?

To produce guidelines that apply in practice, the needs and interests of the target groups had to be considered. Service design is a user-centered way to develop products and services. Even though the goal for this work of thesis is not a concrete product or a service, but rather guidelines for the creation of a digital service, applying some of the methods from service design gives a user-centered aspect. The user-centered insights were considered through the needs of stakeholders.

The interviews and workshops were done as part of the researcher's public officer duties. Being part of the organization provided a possibility to get feedback from actual stakeholders related to the development and use of AI systems. This kind of activity has features of action research, which is an iterative process combining theory and practice and involving researchers and practitioners (Avison et al. 1999, 94).

The qualitative research concept of saturation was applied in the research on collecting the information. The concept of saturation means that information is collected until the need is saturated or until other new materials do not provide new information on the subject. (Dey, Ian, 1993). The research work on publications and regulations culminated on April 21st 2021, when European Commission published the long awaited proposal for the regulation on artificial intelligence.

This research used also interviews of professionals as a way to collect information about artificial systems and the way they are used in the Finnish police. This information received from the interviews was then discussed in co-operative workshops with a constructive perspective. Constructive in a way that the information is gathered from different officials and practices, compared with the proposal for the EU regulation and finally created into a set of principles to be published.

4.2 Service Design approach

A few methods known to service design were used in the study to gather information and refine the outcome with a user-centered and collaborative approach. The methods used were in-depth interviews, peer-reviews and co-creation, collaborative workshops, and prototyping. Stakeholder map was used as a tool to represent different stakeholder groups. (Stickdorn et al. 2018.)

Stakeholder map is a tool used for visualization of different stakeholders for the system. It describes the different stakeholders of the system in different levels (A, B and C) and how they are connected to each other. Stakeholder map can help in reflecting how different stakeholders affect to question at hand. (Stickdorn et al. 2018.)

In-depth interviews are used to gather information about a specific subject from users, experts or stakeholders. Interviews can be semi-structured and conducted face-to-face, telephone or online. According to Steve Portigal in-depth interviews of customers or users can provide crucial shifts in perspective revealing new frames and opportunities that can flip the initial problem on its head. The in-depth interviews can also be used to refine design hypotheses. The interview is better held in the premises or context of the interviewee and the questions should be set the way the interviewee feels as being the expert. The questions

should be designed in advance and consist of background, organizational, projection and wrap-up questions. (Portigal 2013.)

Some techniques to support conducting the interview and help in gaining more valuable information in the interview are participant observation and participatory design. In participant observation the interviewer can ask the interviewee to demonstrate how they use the products or services. Participatory design refers to an overall approach to a design giving the interviewees the tools to envision the product or take part in the development or refinement of the output. (Portigal 2013.)

Peer-review and co-creation in collaborative workshops are good methods to get feedback and reduce risks of confirmation bias early on during the research work by including for example colleagues or stakeholders and inviting them to review the work. Prototyping is used to explore, to evaluate and to communicate. Prototyping is a great way to get quick results on tangible artifacts. Explorative prototypes are used to compare different options and can be just fast drawings on paper. Evaluative prototypes are used to reduce the number of options and give perspective to which to focus on. With communicative prototyping important aspects of the work are communicated to the selected audiences. (Stickdorn et al. 2018.)

Figure 4 shows the methods used for this research in a double diamond. Double diamond describes the different phases of service design process: Discover, Define, Develop and Deliver.

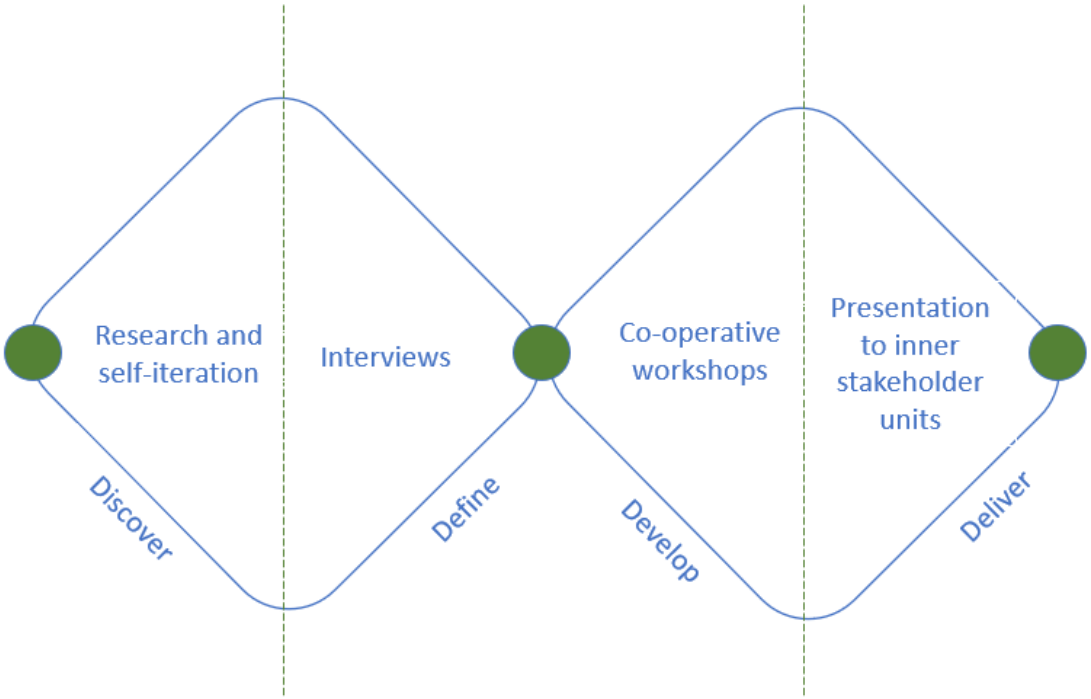


Figure 4. Double diamond with methods used

4.3 Design and implementation of the ethical principles of AI

The aim of the work is to apply service design methods to create the ethical principles of AI for the Finnish police in such a way, that the developers and owners of the systems get concrete benefits of the principles in the development and deployment of AI systems. The principles should be taken as guidelines from the beginning of planning of the systems. The principles should also take into account laws and regulations guiding and restricting the work of the police.

For the principles created by this work of thesis to be taken to use, they should be taken as part of the tasks and requirements of the software development process of the National Police Board. To be taken into practice the principles should be understandable and available at the right place in the right time. To consider the user perspective, the study was carried out by using some applicable methods of service design as they consider the outcome by the needs of customers and end-users. In a way these ethical principles should affect the whole development and implementation process of AI systems.

The development process of the trustworthy principles of AI is described in the Figure 5.

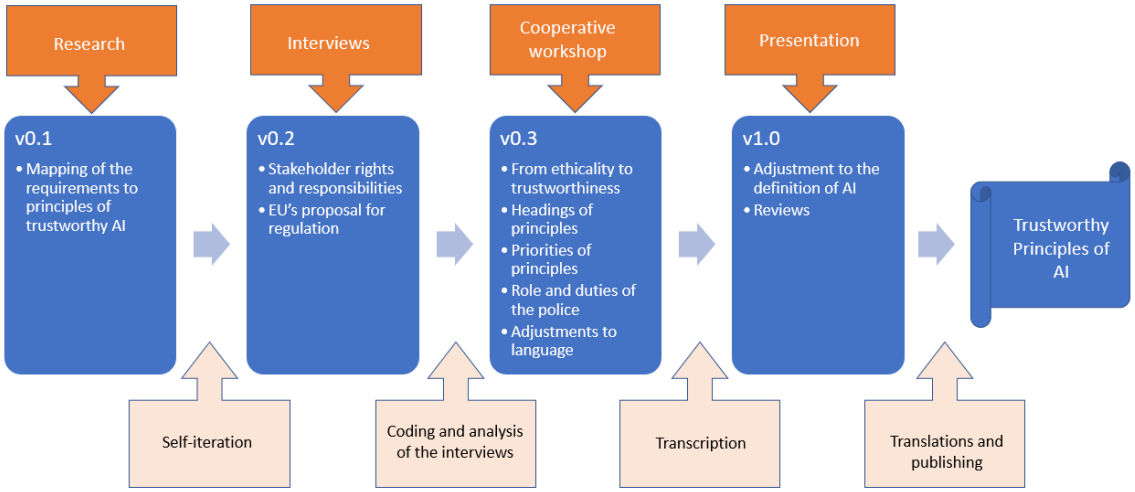


Figure 5. The development process.

The development was done in phases by evolving the version under development. The evolvement was done in different phases by considering different aspects and by using different methods. Each of the phases produced a further developed version of the principles as an outcome.

The version of the principles is shown along with the major changes included in the version in the blue boxes. In the upper orange boxes are shown the major events evolving the development. Interviews and cooperative workshops are methods known to service design. As

the version evolved, the more widely there were participants from the organization. The lower lighter orange boxes include different work phases done in between the phases.

4.4 Starting the work

After extensive study of the subject based on research and literature, the researcher organized an informal kick-off session to begin the creation of the ethical principles of AI for the Finnish police. The kick-off session was held at the end of February 2021 and included a few participants from the National Police Board involved with projects developing AI aided systems. At the session, the researcher told participants about her intention to create ethical principles for artificial intelligence and the potential stakeholders of the project were discussed. The first collaborative meeting around the development of ethical principles of AI for the police was organized in the mid-March 2021. In the collaborative meeting project managers from the projects working with AI systems and two officers from data management unit participated. In the meeting the difference of principles of ethical AI and the concrete norms and instructions of the development of ethical AI was discussed. It was also discussed that the work must be started by aligning the ethical principles of AI and describing the requirements associated with them.

The sketching of ethical principles started by comparing the different principles and guidelines. The EU guidelines of trustworthy AI were considered as the leading guidelines of creation of the principles as they implement the human-centered approach and reflect several known ethical guidelines and methods, like the Public Voice guidelines and the method of Ethically Aligned Design (EAD). The guidelines also form the foundation for the upcoming regulation. Creation of ethical principles of AI started by mapping the ethical principles of trustworthy AI to the requirements of trustworthy AI as described in the figure 6.

Ethical principles and requirements of trustworthy AI

	Respect for human autonomy	Prevention of harm	Fairness	Explicability
Human agency and oversight	x			
Robustness and safety		x		x
Privacy and data governance	x	x	x	
Transparency				x
Diversity, nondiscrimination, and fairness	x		x	
Societal and environmental well-being		x	x	
Accountability			x	x

Figure 6. Mapping of the principles with the requirements.

After the principles of trustworthy AI were mapped with the requirements, the first version, or prototype, of the ethical principles for the police was created by the researcher. This prototype of the ethical principles was drafted by considering the mapped requirements and principles described in the figure 6. The realization of the requirements mapped to each principle is described in the text under each principle. The text describing the realization of the ethical principles also aims to reflect the eight principles of EAD described in the IEEE treatise and the cornerstone principles of Berlin declaration aimed at public sector authorities. The first draft (v0.1) of the ethical principles of AI of the Finnish police is presented in the Table 1.

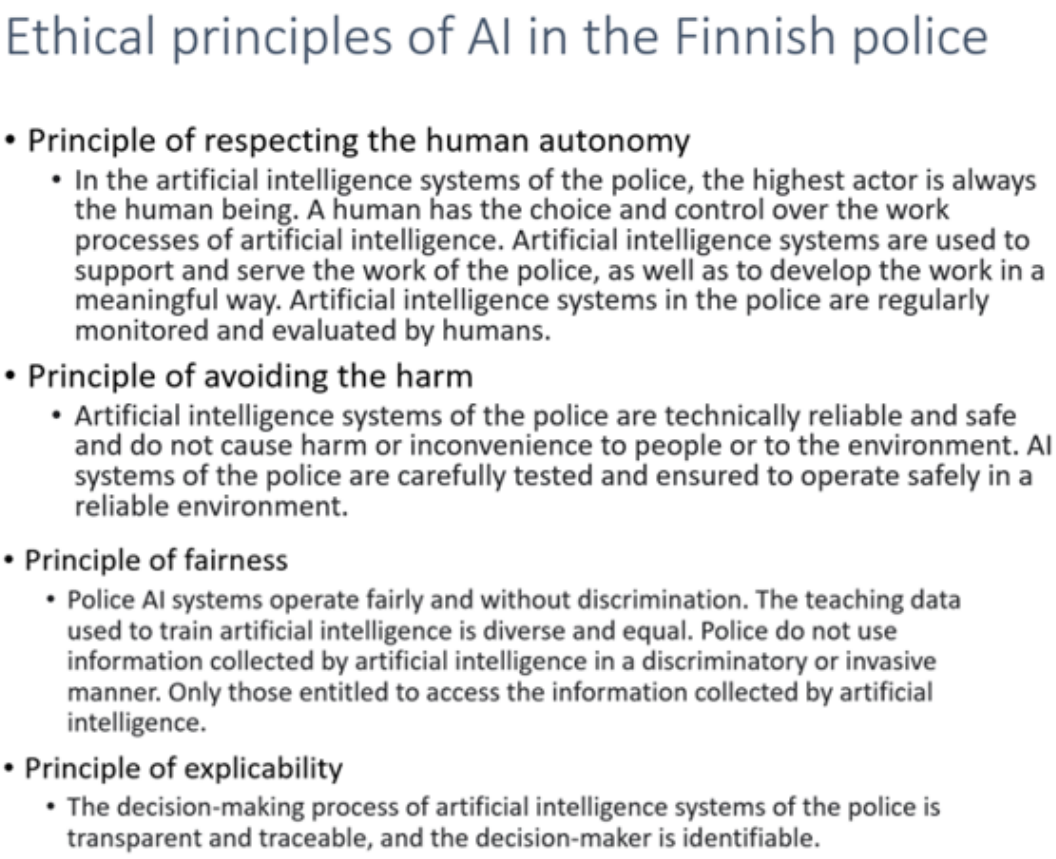


Table 1. First draft (v0.1) of the ethical principles for the Finnish police

4.4.1 Stakeholders of AI systems development

To get a holistic approach to the research problem, the principles in development needed to be considered from the perspective of different stakeholders. The stakeholders of an AI system of the Finnish Police were recognized as described in the stakeholder map in the figure 7. Stakeholder maps are tools of service design describing different stakeholders. With the help of a stakeholder map the needs and requirements of different stakeholders can be categorized.

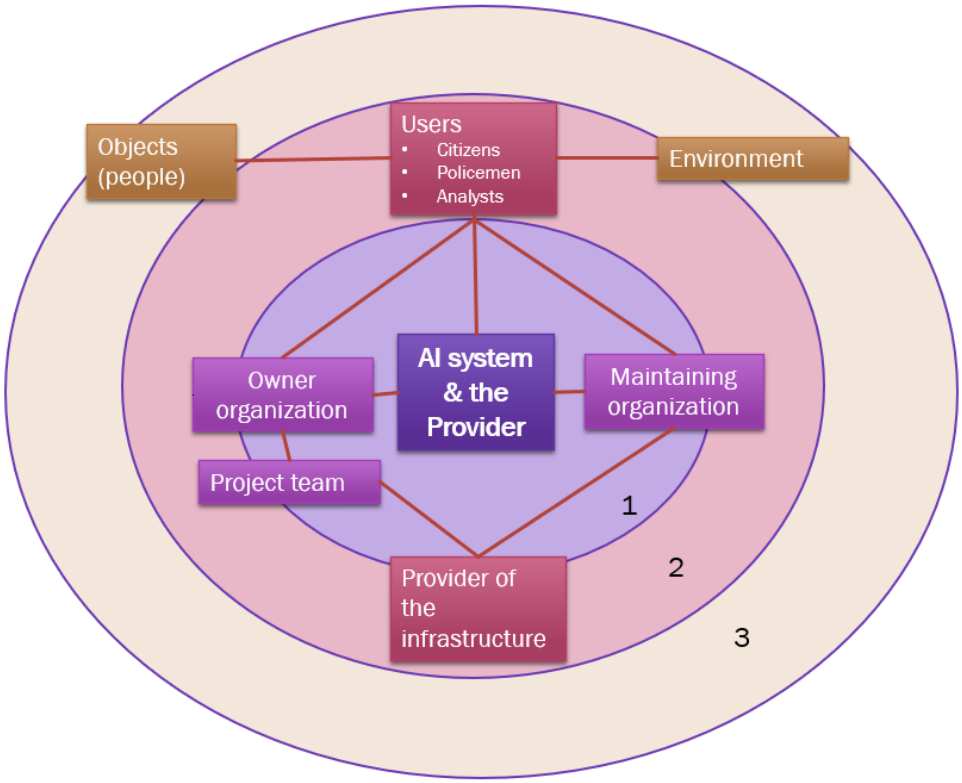


Figure 7. Stakeholder map of an AI system in the National Police Board.

Ethical development and use of AI systems affects all the stakeholders of the systems. The ethical aspects also need to be considered by the objectives, needs and responsibilities of the stakeholders. The principles in development were reflected to consider the stakeholders described in figure 7.

The stakeholders were reflected by the responsibilities and rights of each stakeholder against the requirements of the ethical principles. Users, objects (people) and the environment have rights considering the ethical principles, other stakeholders have only responsibilities, including users (except for citizens) of the system.

The implementation of ethical principles leads eventually to the development of instructions and internal guidelines concerning the development, implementation and use of AI systems. These instructions and guidelines are targeted for the inner use of police.

The inner stakeholders concerned are the ones in the groups of circles 1 and 2. Stakeholders of the inner circle group 1 (purple) are affected with the specification, implementation, and deployment of the system and have responsibilities concerning AI systems. Stakeholders of the middle circle group 2 (rose) can have both responsibilities and rights concerning AI systems.

Group 2's stakeholders can be either users of the system or providers of the infrastructure. They get the system as the way it is but need to follow instructions and processes concerned with it. Inner users of the system, like policemen and investigators, have rights to get instructions, but have the responsibility to use the system as instructed. As for the users outside of police organization, they have only rights concerning the AI system. According to the requirement of transparency, people searching for information on the website and communicating with a chatbot should be made aware that they are communicating with AI.

Stakeholders of the outer circle (group 3) don't have any responsibilities affecting the system, only rights. Below are descriptions of the stakeholders and their responsibilities or rights concerning AI systems.

AI system & the Provider organization is the organization delivering the system for the Finnish police. The provider organization has the responsibility to provide the police an ethically sustainable assessed and documented AI system (within the applicable European and national regulations) as agreed with the Finnish police.

The owner organization is normally a unit of National Police Board or National Bureau of Investigation. The owner organization has the need for the AI system to automate its processes. Therefore, it has the responsibility to provide requirements and steer and guide the development project and ensure the ethical aspects of the system are considered. The owner organization has also the responsibility to gather the dataset needed for the teaching of the AI system. The owner organization is also responsible ensuring the communication of the ethical principles and guidelines for the organizations providing services.

A project team is a team of professionals consisting of project manager(s), functional experts, system administrators, developers, and testers. The objective for a project team is to develop and implement a solution (scope of the project) within the time and budget. The project team has the responsibility to implement and deploy an ethically sustainable system according to the requirements and instructions. The project team has the need for requirements and ethically sustainable dataset for the teaching.

Users are the ones using the system. They have the need for a steadily operating, non-discriminative and achievable system with correct information. They have a right to know if they are communicating or if their issues are handled with an AI system. The inner users (policemen or public officers) of the system are responsible for using the system for ethically sustainable purposes and have the right for documentation and instructions of the system. The end users of the system can be:

- Citizens asking help on searching for information or using of the electronic services.

- Policemen investigating or analyzing crimes.
- Public officers analyzing the data to improve the public services.

Maintaining organization is the organization responsible for maintaining the system in the production. They have a right for documentation and instructions of the system. The maintaining organization has the responsibility to update the system and the documentation, gather and monitor information about the functioning of the system and participate in assessments of the system.

Organization providing the infrastructure for the AI system has the need for requirements of the infrastructure of the system. They have the responsibility to provide a secure, available, and stable user experience for the AI system. They can also have the responsibility to assess the infrastructure provided.

Objects are the people whose information is gathered or used in the AI system or the environment in which the system is working. People as objects of the system have the right to privacy, the right to be treated fairly and without discrimination and the right to be forgotten within the limits of the law. In the crime prevention and investigation systems of the Finnish police people as objects have the right to have their data processed only when it relates to a case. The environment has the right to be treated in a sustainable way.

The responsibilities and needs of the stakeholders were considered in the next version (v0.2) of the ethical principles. The role and rights of users were considered in more detail compared to the previous version.

4.4.2 Requirements of the European Regulation on Artificial Intelligence

The proposal for the European Regulation on Artificial Intelligence was published on April 21st, 2021. The proposal emphasizes risk managerial approach to AI systems and regulates strongly on requirements concerning the high-risk AI systems. (Proposal for a Regulation of the European Parliament and of the Council 2021.) After the publication of the proposal some specifications and additions to the ethical principles of AI for the Finnish police were made to support the implementation of the upcoming regulation. The specifications concerned the principles of prevention of harm, fairness and explicability emphasizing the need of purpose that the use of an AI system is always based on a specific need and an existing purpose for data processing. These additions highlighted in the proposal were also considered in the version 0.2 of the ethical principles. The version 0.2 of the ethical principles is shown in the Appendix 1.

4.5 Verification of suitability of the principles in in-depth interviews

After completing the version 0.2 of the principles found in the Appendix 1, it was time to conduct the in-depth interviews with authorities involved with AI systems in the National Police Board and the National Bureau of Investigation. Altogether seven authorities were interviewed.

The interviewees were asked about their relationship and experience of AI systems of the Finnish police and about their knowledge of ethical issues considering AI systems. Also, the applicability of the principles concerning the present, planned and systems in development was ensured and opinions and requirements for the implementation of the instructions was surveyed. The preliminary questions used in the interviews are found in the Appendix 2. The interviews were held in the end of May 2021.

The interviewees were all somehow familiar with ethical aspects of AI systems, some more, some less. In some recent or ongoing projects, the ethical aspects had been in conversations in the development or in the implementation phases of the project, but there was not any planned process or tool to check the ethical validity of the system.

All interviewees felt that the v0.2 of the ethical principles was broadly in line with the development and use of artificial intelligence systems in the police and that the ethical principles supported the goal of maintaining a high level of public confidence in the police. However, some changes in wording, as well as clarifications to the descriptions, were found necessary to make the principles clear to follow and easy to understand. In almost half of the interviews, it was discussed that it is not just the ethical issues that are considered in the principles but also other aspects of the trustworthy AI, such as logging or instructions. The interviewees also all agreed that for the principles to be truly implemented, their requirements must be taken into practice as part of the normal application development process.

Most of the comments to the ethical principles from the interviewees concerned the quality and ethicality of the dataset used for training the facial recognition system and the education of the of the AI systems (data categorization or visual recognition) to the users. One interviewee emphasized the role of the police as an authority having a duty to work in a fair and non-discriminative way. It means that there is no need to specifically emphasize or disclose the fair and non-discriminative way of working with AI systems as it is already the normal way of working.

Documenting and instructing of the system is part of the normal implementation process. It was discussed that in the training of the AI systems it needs to be emphasized that the AI systems are designed and built to support the police in their work rather than to do the work

for them. AI systems provide support by going through large masses of data that would otherwise take significantly more time by human labor. For example, the facial recognition system does not provide the user any exact search results, but it provides the user a narrow set of results matching most with the image used for the search. The system also shows the probability of match as percentage. This supports the goal that the users of AI systems must be aware that AI systems provide support in searching and analyzing the data, rather than providing the user any exact answers.

The principle of explainability was seen very important in the interviews emphasizing the fact that the operation of the AI system must be predictable. On the other hand, it was also seen somehow hard to implement as the police might not have full descriptions available of algorithms bought from providers. Documentation of the system functions was also seen significant as it supports the situations where the information gathered by the AI system is used as evidence of crime in legal proceedings.

4.6 Analyzing the data collected from the interviews

The in-depth interviews were transcribed and coded. Coding included words like “training data”, “assessments” or “instructions”. The coded words were categorized as “general” or as per principle. Some observations made by the interviewees from the version 0.2 of the principles presented to them were to be presented and discussed later in the workshops. Figure 8 shows how the coded word were categorized.

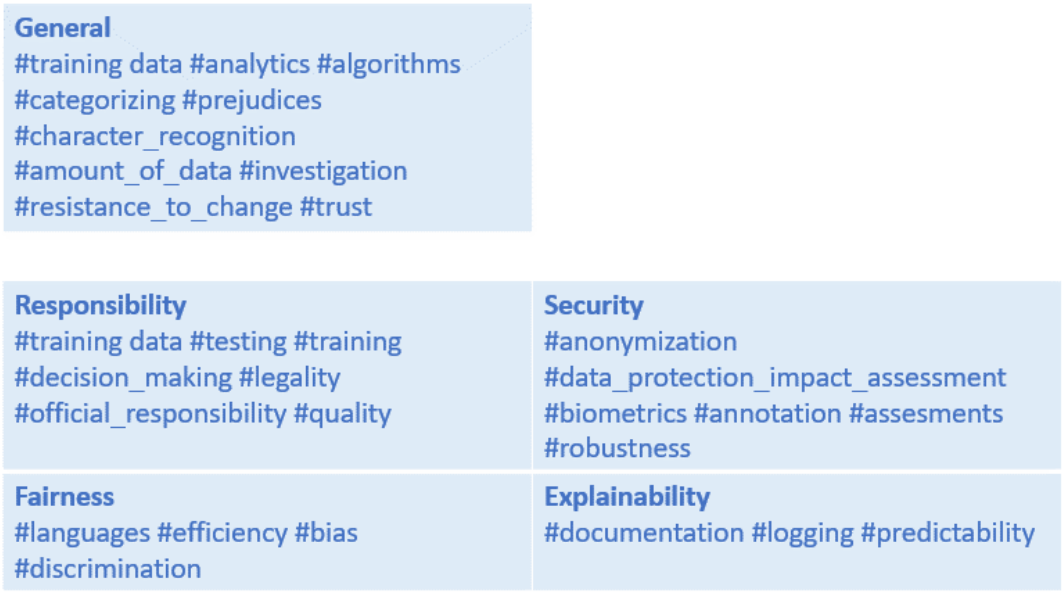


Figure 8. Coding and categorizing of words.

After the interviews were held and discussions were analyzed, some changes were made to the ethical principles. First, the heading “Ethical principles of AI the police” was changed to “Trustworthy principles of AI of the police” to show that besides the ethical aspects the principles are also compliant with law and reflect the requirements of robustness. The names of the principles were changed to simpler versions, for example the “Principle of respecting human autonomy” was changed to “Responsibility” and “Principle of avoiding the harm” to “Security” as shown in the figure 9.

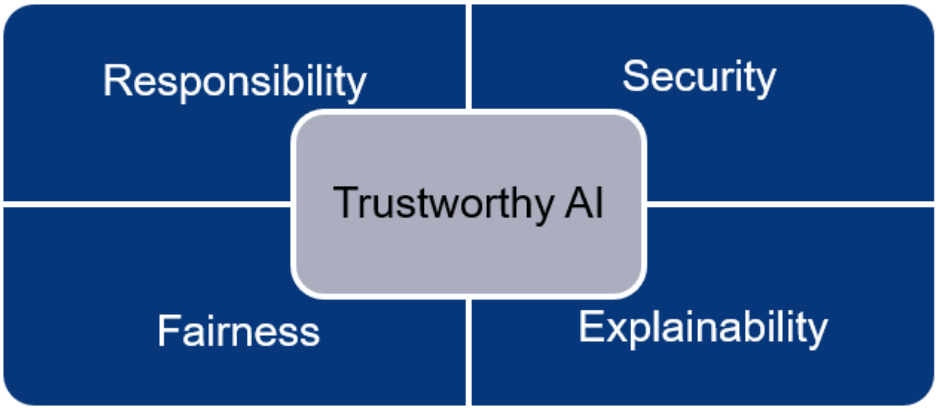


Figure 9. Principles of Trustworthy AI.

Some wordings were also changed to more common language. Also, the role of the police as an authority acting in accordance with the law and regulations was added to the principle of fairness. This version of the principles was used as the working document in the cooperative workshop.

4.7 Working in cooperative workshops

After analyzing the information received in the interviews and adjusting them to the principles, a cooperative workshop lasting half of a day was organized. The participants consisted of seven public servants, of which 5 were different from the previously interviewed. The workshop was held in June 2021. All participants were familiar the concept of artificial intelligence and some even with the ethical issues of AI. The participants were provided with the version v0.2 of the principles for prior reference. To start the workshop the researcher introduced the subject of ethicality of artificial intelligence. The attendees were shown the version with adjustments from the interviews along with issues concerning the coded words to raise the discussion.

The discussion in the workshop over the trustworthy principles was very enthusiastic and fruitful as the attendees had many good remarks on the subject.

The attendees of the workshop felt that as the target group for the principles is citizens it is necessary in the beginning to clarify what is meant with artificial intelligence in the Finnish police. A brief description of AI and how it is used in the police was added to the beginning of the principles. Also, a description of the role of the principles was added to the first chapter.

The attendees also felt that as the Finnish police enjoys a high level of trust among the citizens and the situation is to be maintained, the role of official responsibility needs to be emphasized showing that it is the police officer that has the responsibility over the use of AI systems and AI is used to support the duties of the police officers in an effective and appropriate way. The principle of "Responsibility" also highlights the fact that the AI systems of the police are built to support fundamental rights and the ethical use of AI; e.g. facial recognition system results with multiple hits telling the user the percentage of probability of each hit.

The principle of "Security" deals mostly with reliable and secure AI solutions, but also with training the users of the system. A sentence about utilizing AI only for processing purposes in accordance with the regulations in force was added to support the fact that to support the implementation of fundamental rights the use of the AI systems is controlled, e.g., the use of facial recognition system is only allowed if the punishment for the offense in question were imprisonment. The role of training the system and analyzing the results received with the aid of an AI system was emphasized as it was one of the requirements for high-risk systems in the EU's proposal for the regulation for AI systems.

The description of the principle of "Fairness" emphasizes the duty of the police to act in accordance with the regulations. The attendees of the co-operative workshop also wanted to highlight the roles of accessibility and availability as means of providing fair and equal transaction services to citizens regardless of time or place. The principle of fairness also considers the ethicality of the training material used for training AI systems. In the workshop it was discussed whether the requirements of ethicality should be opened in the text, but everyone agreed that the requirements considering the training material are better to be described at a higher level and the more detailed requirements are introduced in the internal instructions of the organization implementing the AI system. The implementation of privacy protection was also emphasized in the text both in the technical and user level.

The descriptive text of the principle of "Explainability" was tuned to clearer format and a sentence about documentation supporting the use of information gathered and analyzed by AI systems in police operations was added.

The text describing the trustworthy principles was also edited from a passive spelling to the first form of the plural to fit the form of text describing the ethical codes of the police found

on the website of the Finnish police. The version 0.3 of the principles found in the Appendix 3 is the outcome of the cooperative workshop.

4.8 Presentation of the principles in development

After the principles were transcribed from the changes considered in the cooperative workshop, the next step was to introduce the principles to the authorities representing different units of the National Police Board and National Bureau of Investigation. A presentation session for staff of the draft of the trustworthy principles was held in the end of August 2021. The background material, methods of working, the trustworthy principles themself and factors behind the principles were described and introduced to the participants. The participants consisted of representatives from each unit of the National Police Board and from the authorities participated in specifying of the principles.

Feedback received in the session was supportive, though some adjustments were considered necessary. The definition of artificial intelligence and the term of AI system were considered to need some clarification in the version 0.3 of the trustworthy principles of AI. Defining ownership of principles in the organization and incorporating the principles into the data protection impact assessment also sparked conversation among participants.

After some research and conversations with a couple of specialists the definition of AI and some clarifications of a few terms were adjusted to the principles. The imitation of the human activities and the interaction with the environment were added to the definition of AI.

The last steps of the process were to introduce the principles to a steering group of digital development and publish the principles of trustworthy AI in the www.poliisi.fi in Finnish and Swedish.

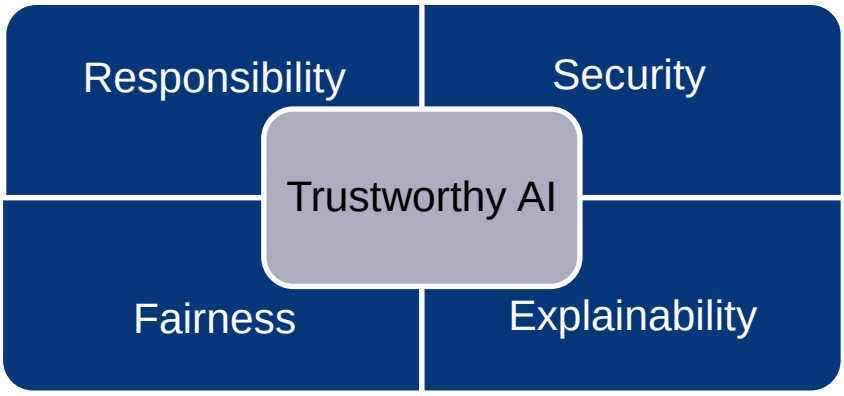
4.9 Final version of the trustworthy principles of AI

The trustworthy principles of AI for the Finnish Police were developed in phases described above during the period of May and October 2021. The development was done in collaboration between several units of the Finnish National Police Board and The Bureau of Investigation. The final version of the trustworthy principles of AI is presented above.

> **The Trustworthy Principles of Artificial Intelligence**
>
> By artificial intelligence, we in the police mean a system that interacts with its environment by identifying, observing, and analyzing it, and producing results that reproduce human thinking based on this action. Artificial intelligence does not replace a police officer as an actor or a decision maker.

In the police, we use artificial intelligence systems to support and enhance our operations in our statutory duties, as well as to achieve our goal of being a security guard for all, at all times. We utilize artificial intelligence responsibly and the use of artificial intelligence is always controlled by man.

The principles of reliable artificial intelligence described here guide the lawful and ethical application of artificial intelligence in the police.



Responsibility

Artificial intelligence is used to enhance and serve the work of the police in an appropriate way.

The highest authority and responsibility for systems utilizing artificial intelligence is always with the police officer, who makes choices and decisions based on information obtained with the support of artificial intelligence. The use of systems utilizing artificial intelligence and the interpretation of outputs have been trained and instructed for our users.

In the police, we actively monitor the development of artificial intelligence in relation to our own operations and develop our skills and operations accordingly.

Security

The artificial intelligence systems we use are technically reliable and secure. The use of artificial intelligence does not cause harm or inconvenience to humans or the environment.

Prior to the introduction of a system utilizing artificial intelligence, impact and risk assessments of the system are performed, and risk management measures are planned.

We regularly test, monitor, and assess systems utilizing artificial intelligence.

Fairness

Artificial intelligence systems introduced in the police are implemented in an ethical manner, as well as to respect fundamental rights such as privacy and equality.

We use artificial intelligence systems to increase the efficiency of our operations, as well as to provide citizens with time and place-independent business opportunities.

The use of artificial intelligence in the police supports equality, is always in accordance with the law and regulations and is used for processing purposes in accordance with the regulations in force.

The material used to train artificial intelligence is appropriate, of high quality and meets ethical requirements. The artificial intelligence systems we use are ensured to operate in a privacy-friendly and non-discriminatory manner.

Explainability

Artificial intelligence systems of the police operate regularly and predictably. The operation of the artificial intelligence system is traceable and verifiable, and the actor responsible for decision-making is always identifiable.

With the help of documentation, we support the utilization of information produced and analyzed with the help of artificial intelligence in our operations and decision-making.

5 Discussion

The Finnish police works and operates according to laws, regulations, and the ethical codes of the police, one could ask “What is the added value of these principles?”. The trustworthy principles of AI are an extension to laws and regulations and on top of the regulative approach they express the ethical commitment to use AI to provide beneficence. EU has only recently published the proposal for the regulation concerning AI systems, so it will take some time before the regulation is binding. Yet there is pressure to take advantage of AI as the digital information is expanding rapidly. The most important function of the trustworthy principles of artificial intelligence is to promote trust in the way the police and the public sector use artificial intelligence in their operations.

The organization of the police is a formal organization following many different processes. The trustworthy principles of AI need to be integrated into the information systems development process and to the actual usage procedures of the AI systems. The principles need to be published and made aware in the organization to be truly internalized and taken into practice. Information about new practices and instructions should be communicated internally within the organization to those responsible for the development of artificial intelligence software, as well as to users of AI software. The requirements of the trustworthy AI need to be modified into tasks or checkpoints in the application development process. A virtual self-paced training on the principles of the trustworthy artificial intelligence in practical work would also support the goal of implementing the principles into practice.

As there were no existing organizational structures concerning AI systems in the National Police Board and the few existing implementations of AI were scattered throughout the organization, service design methods were a good way to collect information and views on the ethical application of artificial intelligence. The co-operative workshop also served as a first step or pilot for the joint AI working group in the organization. It is clear the trustworthy principles of AI are just a foundation for a broader methodological development on requirements concerning AI systems. More detailed instructions and procedures on implementing these principles are needed to guide the development, implementation and use of AI systems in the Finnish police.

The trustworthy principles of AI of the Finnish police describe the requirements for the development, implementation and use of AI systems in a high level. Each requirement of EU's trustworthy AI is included in the text describing the principles to explain the public how the trustworthiness and ethicality is implemented in the police. As there are different kinds of AI systems and probably there are more to come, the principles need to be described in high-level to avoid any unnecessary commitments or frequent updates to the principles. When designing or implementing a new AI aided system, the requirements of the principles need to be considered against the system in question. The practical applicability of the principles of trustworthy AI should be reviewed periodically and updated if deficiencies or needs for change are identified.

Most of the publications studied for this thesis dealt with the ethics of AI, but to be precise AI does not have ethics in it, the issue concerns the ethical application of artificial intelligence. AI systems learn from what they are taught and how they are used, and that is why deep judgement and planning needs to be carried out in designing, implementing, and instructing of the AI systems and their training material.

A question to consider is also that in which kind of situations and operations the use of AI systems (e.g., facial recognition system) is acceptable when the AI system affects the

fundamental rights of the citizens. The Finnish police has instructed the use of automatic facial recognition system is permitted only in criminal investigation situations where the offense is punishable by imprisonment. The upcoming EU regulation considers real-time facial recognition systems banned except for couple of exceptions (e.g., missing child, suspect of a terrorist attack). Considering the principle of "beneficence", would it be unethical to use AI systems to analyze masses of data and help the police to solve also other, less serious, crimes? Is it ethical to not use all material available? Does the fear of the control society overcome the seizure of the benefits of artificial intelligence? Can the trust of citizens and society be built and foundations layed even for a social artificial intelligence ecosystem by communicating and committing to the ethical principles of artificial intelligence?

6 Conclusion

The aim of this thesis was to develop ethical principles for the design, implementation and use of AI systems. The first draft of the principles was written after familiarizing with some known guidelines of AI ethics and EU's proposal for regulation on AI. The work continued using methods known to service design, like in-depth interviews, co-operative workshops, and prototyping. These methods were used to guarantee that the principles in development fit the needs and requirements of the Finnish police.

The applicability and trustworthiness of the principles developed in this work of thesis was confirmed by inviting some authorities involved with AI systems to the interviews, co-operative workshops and to the final presentation session. The authorities represented directorial, managerial, and operative levels of the National Police Board and The National Bureau of Investigation. By involving different stakeholders, a broader viewpoint was created. This thesis has shown that service design methods suit well in creating of textual communications material of law enforcement duties.

The use of the service design methods, in-depth interviews, and co-operative workshops allowed a review to the needs and perspectives of distinct organizations and experts in the design of the ethical principles of artificial intelligence. Building the principles by prototyping made the principles tangible and served as a basis for discussion and further development. Developing the principles by involving professionals from different units in collaborational workshops and in in-depth interviews also laid a horizontal perspective to the principles.

The use of the service design methods, including drafts and prototypes, seemed also a good way of working on a reasonably new topic. A tangible product, or a draft of a document in this case, set out the substance of the matter, which was adapted into an understandable form describing some operational processes involved with AI systems.

This work of thesis provided the Finnish police a set of trustworthy principles to guide the development, implementation and use of AI systems. By taking the principles into practice, the Finnish police can strengthen the trust among the citizens by showing them the interest of acting as pioneers of law enforcement and considering ethical issues when working with AI systems. By communicating about the principles inside the organization, the overall knowledge of the subject is increased which will affect the ethical application of AI systems.

The external validity of these trustworthy principles can be assessed by looking at how they are implemented in practical work, both in the development of artificial intelligence systems and in their use. The aim was also to increase awareness in the organization regarding the ethical application of artificial intelligence, and the realization of this goal can be examined by finding out the staff's awareness of the topic. These trustworthy principles of AI are designed to guide the authorities of the Finnish police in the development, implementation and use of AI systems to take special attention and consideration on activities affecting fundamental human rights. The principles developed are suitable also to other law enforcement authorities with minor changes.

The trustworthy principles of AI provide changes to organizational processes when new tasks are incorporated into the processes. The change management process is an ongoing process. Once the change is implemented to organization, the realization needs to be reviewed and new actions to support the change identified, analyzed, and planned. The figure 10 presents the change management process of an organizational change, like the application of the trustworthy principles of AI. After the principles are published and communicated, the application of the principles in practice can be reviewed by a survey and by assessments. If changes to the principles or their implementation is needed, the change must be analyzed planned and implemented. After the implementation of the change, the application of the principles is to be reviewed again and further changes analyzed.

Figure 10. The change management process.

The research material used for this thesis of work is all public. In accordance with the guidelines and research permit of the National Police Board no names of the authorities taking part in the interviews or co-operative workshops, nor material gathered in them is published.

Topics for further development involve the implementation of the inner instructions of the trustworthy principles, as well as analyzing the realization of the principles in future AI systems development of the Finnish police. It would also be interesting to do research on the design and implementation of other public authorities' ethical principles or about the preparation on the upcoming EU regulation and how it affects different kind of public authorities.

7 Appendices

Appendix 1: Version 0.2 of the ethical principles

Ethical principles of AI in the Finnish police

- Principle of respecting the human autonomy
  - In the artificial intelligence systems of the police, the highest actor is always the human being. A human has the choice and control over the work processes of artificial intelligence. Artificial intelligence systems are used to support and serve the work of the police, as well as to develop the work in a meaningful way. Artificial intelligence systems in the police are regularly monitored and evaluated by humans.
- Principle of avoiding the harm
  - Artificial intelligence systems of the police are technically reliable and safe and do not cause harm or inconvenience to people or to the environment. AI systems of the police are carefully tested and ensured to operate safely in a reliable environment. AI systems of the police always include instructions, and the users are trained to use them before deployment.
- Principle of fairness
  - The use of an AI system is always based on a purpose. Police AI systems operate fairly and without discrimination. The training data used to train artificial intelligence is relevant, presentative, diverse and equal. The police does not use information collected by artificial intelligence in a discriminatory or invasive manner. Only those entitled to access the information collected by artificial intelligence.
- Principle of explicability
  - The decision-making process and the functioning of artificial intelligence systems of the police is predictable, transparent and traceable, and the decision-maker is identifiable. Users of the system understand the functioning of the system and they understand they’re acting with an AI system.

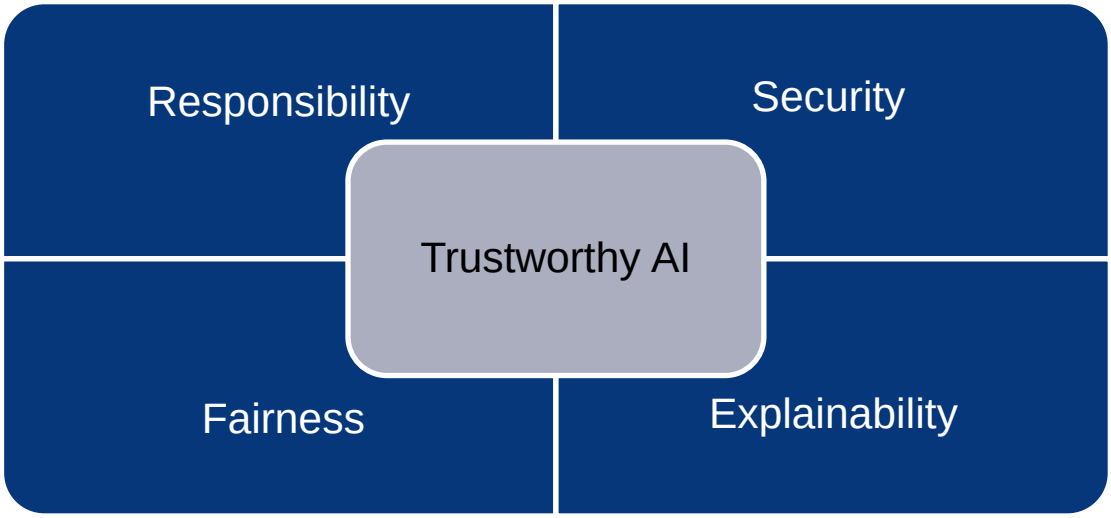
Appendix 2: Questions of the interviews

Designing of the Ethical Principles for Artificial Intelligence for the Finnish Police	
Laurea University of Applied Sciences / Leading Transformational Change	
Questions	
1.	***Name and organization in the police? Relationship or interest towards AI systems and their development?***
2.	***Are there AI systems used or "owned" by your organization? Or are there plans to introduce one?***
3.	***If yes, what kind of AI systems? For example:*** • ***Systems concerned with speech or text literation.*** • ***Systems concerned with image recognition*** • ***Systems concerned with data classification*** • ***Something else, what?***
4.	***Are you aware if the system is categorized as high-risk AI system?***
5.	***Are you aware of any guidelines concerning the ethical aspects of AI systems? For example: "EU Guidelines for Trustworthy AI" or "Public Voice Guidelines" or "IEEE standard for Ethically Aligned Design"***
6.	***Were ethical values a criterion when purchasing the system?***
7.	***Have the ethical requirements of the system been discussed with the supplier when purchasing the system?***
8.	***Has the ethics of AI teaching data been discussed with the supplier or internally in connection with the teaching and implementation of the system?***
9.	***Has the supplying organization of the AI system published it's ethical principles of AI and is it committed to them?***
10.	***Has the information management organization been involved with the development and implementation project of the AI system?***
11.	***Are the requirements of the new personal data law been taken to account in the development and implementation of the AI system?***
12.	***Who (organization) is responsible for teaching of the AI system and acquiring of the dataset for teaching?***
13.	***How is the technical monitoring of the AI system conducted?***
14.	***How are the decisions made by the AI system monitored?***
15.	***How is the AI system documented and how is the documentation maintained?***
16.	***How do you assess the benefits of using an artificial intelligence system?***

17.	***How do you assess the possible risks of using an artificial intelligence system?***
18.	***Would you be interested in piloting and give feedback about the instructions concerning the ethical principles of AI?***

Appendix 3: Version 0.3. (The trustworthy principles of AI for the Finnish police)

The police use's artificial intelligence-assisted information systems to support and enhance the police's performance in their statutory tasks. Artificial intelligence in police refers to an information system that learns on the basis of activities and that identifies, perceives and analyzes information and the environment. Artificial intelligence does not replace the police as an actor or decision-maker. The use of artificial intelligence is always controlled by a police officer. The principles of trustworthy AI described here guide the lawful and ethical application of artificial intelligence in the police.



Responsibility

The police control the activities of artificial intelligence. The highest authority and responsibility for systems utilizing artificial intelligence in the police is always with the police officer, who makes choices and decisions based on the information obtained with the support of artificial intelligence. The artificial intelligence capabilities of the police are implemented to be guided ethically, as well as to respect fundamental rights. Artificial intelligence is used to enhance and serve the work of the police in an appropriate and effective way.

Security

The artificial intelligence capabilities used by the police are technically reliable and secure. The use of artificial intelligence does not cause harm or inconvenience to humans or the environment. Artificial intelligence is utilized for purposes in accordance with the regulations in force, and an impact assessment is prepared before it is introduced to production. Police information systems utilizing artificial intelligence are regularly tested, monitored, and evaluated. The use, application, and interpretation of the outputs of artificial intelligence are trained for users of artificial intelligence and the use of systems utilizing artificial intelligence is always instructed.

Fairness

The use of artificial intelligence in the police is always in accordance with the law and regulations, and it is used e.g. to provide accessible, time- and place-independent access to citizens to the services of the police. The material used to train artificial intelligence is appropriate, of high quality and meets ethical requirements. The artificial intelligence used by the police has been ensured to work to implement the protection of privacy, and the persons who use and direct the artificial intelligence in the police understand the possibilities and limitations of artificial intelligence.

Explainability

The artificial intelligence of the police operates in a regular and predictable manner. Activities are traceable and the officer responsible for decision-making is always identifiable. Documentation supports the utilization of information produced and analyzed by artificial intelligence in police operations and decision-making.

References

Electronic

Ailisto, H. *et al.* (2018) 'Tekoälyn kokonaiskuva ja osaamiskartoitus', p. 70.

Algorithms and human rights - Study on the human rights dimensions of automated data processing techniques and possible regulatory implications (2018) *COE*. Available at: https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html (Accessed: 27 March 2021).

AlgorithmWatch (2019) 'AI Ethics Guidelines Global Inventory', *AlgorithmWatch*, 9 April. Available at: https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory (Accessed: 27 March 2021).

AlgorithmWatch (2020) 'The only way to hold Facebook, Google and others accountable: More access to platform data', *AlgorithmWatch*, 26 May. Available at: https://algorithmwatch.org/en/governing-platforms-studies-may-2020 (Accessed: 27 March 2021).

Automaattisen päätöksenteon sääntelyä valmisteleva työryhmä asetettu (2021) *Oikeusministeriö*. Available at: https://oikeusministerio.fi/-/automaattisen-paatoksenteon-saantelya-valmisteleva-tyoryhma-asetettu (Accessed: 20 March 2021).

Avison, D.E. *et al.* (1999) 'Action research', *Communications of the ACM*, 42(1), pp. 94–97. doi:10.1145/291469.291479.

Babuta, A., Oswald, M. and Rinik, C. (2018) 'Machine Learning Algorithms and Police Decision-Making - Legal, Ethical and Regulatory Challenges', p. 45.

'Ban Facial Recognition – thepublicvoice.org' (2019), October. Available at: https://thepublicvoice.org/ban-facial-recognition/ (Accessed: 29 March 2021).

'Berlin Declaration on Digital Society and Value-Based Digital Government' (2020). EU. Available at: https://ec.europa.eu/isa2/sites/isa/files/cdr_20201207_eu2020_berlin_declaration_on_digital_society_and_value-based_digital_government_.pdf.

Charter of Fundamental Rights (2020). Citizensinformation.ie. Available at: https://www.citizensinformation.ie/en/government_in_ireland/european_government/eu_law/charter_of_fundamental_rights.html# (Accessed: 29 March 2021).

Charter of Fundamental Rights of the European Union (2012). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012P/TXT&from=EN (Accessed: 29 March 2021).

Council of Europe (2018) 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment'. Available at: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (Accessed: 28 March 2021).

Council of Europe (no date) *CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment, European Commission for the Efficiency of Justice (CEPEJ)*. Available at: https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment (Accessed: 20 March 2021).

Data protection under GDPR (2021) *Your Europe*. Available at: https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_en.htm (Accessed: 21 October 2021).

Dey, Ian (1993) *Qualitative Data Analysis - A User Friendly Guide for Social Scientists*. Routledge.

Dignum, V. (2018) 'Ethics in artificial intelligence: introduction to the special issue', *Ethics and Information Technology*, 20(1), pp. 1–3. doi:10.1007/s10676-018-9450-z.

Eettinen koodisto (no date) *Poliisi*. Available at: https://poliisi.fi/eettinen-koodisto (Accessed: 21 March 2021).

Ethics guidelines for trustworthy AI (2019) *Shaping Europe's digital future - European Commission*.

Ethics of AI (2021). Available at: https://ethics-of-ai.mooc.fi/ (Accessed: 10 April 2021).

European Commission (2020) *EU Member States sign the Berlin Declaration on Digital Society*, *ISA² - European Commission*. Available at: https://ec.europa.eu/isa2/news/eu-member-states-sign-berlin-declaration-digital-society_en (Accessed: 10 April 2021).

European Commission (2021) *A European approach to Artificial intelligence | Shaping Europe's digital future*, *https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence*. Available at: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence (Accessed: 21 March 2021).

FINLEX ® - Säädökset alkuperäisinä: Laki henkilötietojen käsittelystä… 616/2019 (2019). Oikeusministeriö. Available at: https://www.finlex.fi/fi/laki/alkup/2019/20190616#Pidp448652000 (Accessed: 20 March 2021).

Floridi, L. *et al.* (2018) 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations', *Minds and Machines*, 28(4), pp. 689–707. doi:10.1007/s11023-018-9482-5.

'Fundamental rights review of EU data collection instruments and programmes - FINAL REPORT' (2019). FGB, Fondazione Giacomo Brodolini. Available at: http://www.fondazionebrodolini.it/sites/default/files/final_report_0.pdf.

Fuster, D.G.G. and Brussel, V.U. (2020) 'Artificial Intelligence and Law Enforcement - Impact on Fundamental Rights', p. 92.

Ginsberg, M. (2012) *Essentials of Artificial Intelligence*. Newnes.

'Guidelines 3/2019 on processing of personal data through video devices' (2020). The European Data Protection Board.

Hendler, J., Tate, A. and Drummond, M. (1990) 'AI planning: Systems and techniques', *AI Magazine*, 11(2).

Holst, A. (2021) *Total data volume worldwide 2010-2025*, *Statista*. Available at: https://www.statista.com/statistics/871513/worldwide-data-created/ (Accessed: 31 October 2021).

IEEE (2019) ‘Ethically aligned design’. IEEE. Available at: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf (Accessed: 2 April 2021).

Institutions and bodies (2021) *European Union*. Available at: https://europa.eu/european-union/about-eu/institutions-bodies_en (Accessed: 22 March 2021).

Jobin, A., Ienca, M. and Vayena, E. (2019) *Artificial Intelligence: the global landscape of ethics guidelines*.

Kayid, A. (2020) *The role of Artificial Intelligence in future technology*. doi:10.13140/RG.2.2.12799.23201.

Labudzki, R., Legutko, S. and Raos, P. (2014) ‘The essence and applications of machine vision’, *Tehnicki Vjesnik*, 21, pp. 903–909.

Latonero, M. (2018) ‘Governing Artificial Intelligence’:, p. 38.

‘Luotettavaa tekoälyä koskevat eettiset ohjeet’ (2019). European Commission. Available at: https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_FI.pdf (Accessed: 11 March 2021).

McNamara, A., Smith, J. and Murphy-Hill, E. (2018) ‘Does ACM’s code of ethics change ethical decision making in software development?’, in *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ESEC/FSE ’18: 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, Lake Buena Vista FL USA: ACM, pp. 729–733. doi:10.1145/3236024.3264833.

Nadler, D.A. and Tushman, M.L. (1990) ‘Beyond the Charismatic Leader: Leadership and Organizational Change’, *California Management Review*, 32(2), pp. 77–97. doi:10.2307/41166606.

Ollila, M.-R. (2019) ‘Tekoälyn etiikkaa’, *Otava*, March. Available at: https://otava.fi/kirjat/tekoalyn-etiikkaa/ (Accessed: 3 April 2021).

Pfeiffer, R.S. and Forsberg, R.P. (1993) *Ethics on the Job: Cases and Strategies*. Wadsworth Publishing Company.

Poliisi on saanut rikollisia kiinni kasvoja tunnistavan tekoälyn avulla ja haluaisi laajentaa valtuuksiaan – testasimme, miten kone toimii (2020) *Yle Uutiset*. Available at: https://yle.fi/uutiset/3-11448002 (Accessed: 25 April 2021).

Poliisin ja Rajavartiolaitoksen henkilötietolait ja matkustajarekisteridirektiivin täytäntöönpanolaki voimaan 1.6. (2019) *Sisäministeriö*. Available at: https://intermin.fi/-/poliisin-ja-rajavartiolaitoksen-henkilotietolait-ja-matkustajarekisteridirektiivin-taytantoonpanolaki-voimaan-1-6-?languageId=en_US (Accessed: 21 March 2021).

Portigal, S. (2013) *Interviewing Users: How to Uncover Compelling Insights*. Sebastopol, UNITED STATES: Rosenfeld Media. Available at: http://ebookcentral.proquest.com/lib/laurea/detail.action?docID=1173171 (Accessed: 30 April 2021).

‘Proposal for a Regulation of the European Parliament and of the Council’ (2021). European Commission. Available at: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence (Accessed: 24 April 2021).

Regulations, Directives and other acts (2016) *European Union*. Available at: https://europa.eu/european-union/law/legal-acts_en (Accessed: 10 May 2021).

Regulatory framework on AI | Shaping Europe's digital future (2021). Available at: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (Accessed: 24 April 2021).

Rekisterikilven lukulaitteesta tuli kolmessa vuodessa poliisin paras kaveri – "Kaikki kehuvat" (2017) *Yle Uutiset*. Available at: https://yle.fi/uutiset/3-9658191 (Accessed: 29 June 2021).

Rijckaert, M.J., Debroey, V. and Bogaerts, W. (1988) 'Expert Systems: The State of the Art', in Mitra, G. et al. (eds) *Mathematical Models for Decision Support*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 487–517. doi:10.1007/978-3-642-83555-1_29.

Ronsin, X. and Lampos, V. (2018) 'In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data, Appendix I accompanying the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment, adopted by the CEPEJ during its 31st Plenary meeting (Strasbourg, 3-4 December 2018)', in. Strasbourg: CEPEJ. Available at: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (Accessed: 28 March 2021).

Russell, S. and Norvig, P. (2016) *Artificial intelligence : a modern approach*. Third edition. Global edition. Pearson Education Limited (Prentice Hall Series in Artificial Intelligence).

Searle, J. (1980) 'Minds, brains and programs'. Available at: http://et3r.ektf.hu/wp-content/uploads/2014/06/minds-brains-and-programs.pdf (Accessed: 6 May 2021).

Some Observations on the Clearview AI Facial Recognition System – From Someone Who Has Actually Used It … (2020) *CPO Magazine*. Available at: https://www.cpomagazine.com/data-privacy/some-observations-on-the-clearview-ai-facial-recognition-system-from-someone-who-has-actually-used-it/ (Accessed: 28 March 2021).

Statewatch (2019) *Statewatch | EU: PNR: Council wants to 'explore the necessity and feasibility' of the surveillance and profiling of all forms of mass transport*. Available at: https://www.statewatch.org/news/2019/october/eu-pnr-council-wants-to-explore-the-necessity-and-feasibility-of-the-surveillance-and-profiling-of-all-forms-of-mass-transport/ (Accessed: 28 March 2021).

Stickdorn, M. *et al.* (2018) *This Is Service Design Doing: Applying Service Design Thinking in the Real World*. O'Reilly Media, Inc.

The Public Voice (2018) 'AI Universal Guidelines – thepublicvoice.org', 23 October. Available at: https://thepublicvoice.org/ai-universal-guidelines/ (Accessed: 27 March 2021).

Tietosuoja | Krp kokeili kasvojentunnistussovelluksen käyttämistä, testauksesta tehty ilmoitus tietosuojavaltuutetulle (2021) *Helsingin Sanomat*. Available at: https://www.hs.fi/kotimaa/art-2000007911042.html (Accessed: 25 April 2021).

United Nations (1948) *Universal Declaration of Human Rights*, *United Nations*. United Nations. Available at: https://www.un.org/en/about-us/universal-declaration-of-human-rights (Accessed: 2 April 2021).

Usein kysytyt kysymykset poliisin henkilötietojen käsittelyä koskevasta lainsäädännöstä (no date) *Sisäministeriö*. Available at: https://intermin.fi/poliisiasiat/toimijat-ja-vastuut/usein-kysytyt-kysymykset-poliisin-henkilotietojen-kasittelya-koskevista-saadoksista (Accessed: 21 March 2021).

Vakkuri, V., Kemell, K.-K. and Abrahamsson, P. (2019) 'Ethically Aligned Design: An Empirical Evaluation of the RESOLVEDD-Strategy in Software and Systems Development Context', in *2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. *2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, Kallithea-Chalkidiki, Greece: IEEE, pp. 46–50. doi:10.1109/SEAA.2019.00015.

Vuoden 2020 rahoitushaku (2020) *Valtiovarainministeriö*. Available at: https://vm.fi/robohaku2020 (Accessed: 10 April 2021).

Vuorensyrjä, M. and Rauta, J. (2020) *Poliisibarometri 2020 : Kansalaisten arviot poliisin toiminnasta ja Suomen sisäisen turvallisuuden tilasta*. sisäministeriö. Available at: https://julkaisut.valtioneuvosto.fi/handle/10024/162345 (Accessed: 3 October 2021).

'White Paper On Artificial Intelligence - A European approach to excellence and trust' (2020). European Commission. Available at: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (Accessed: 21 March 2021).

Wu, F. *et al*. (2020) 'Towards a new generation of artificial intelligence in China', *Nature Machine Intelligence*, 2(6), pp. 312–316. doi:10.1038/s42256-020-0183-4.

Figures

Tables