


Ethical behavior with artificial intelligence in the ICT-industry: Towards an improved code of ethics

Arnoud Leonard Eduard Nederpel

Faculty of Technology, Policy and Management

NLdigital 

This page is intentionally left blank.

Ethical behavior with artificial intelligence in the information and communications technology industry in the Netherlands

A mixed-method descriptive study of the ethical design and application of artificial intelligence systems in the Netherlands and improvement of the current code of ethics for artificial intelligence in the ICT-sector in the Netherlands

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in Management of Technology

By

Arnoud Leonard Eduard (A.L.E.) Nederpel

Student number: 4303342

to be defended publicly on Thursday, March 5, 2020, at 13:45 PM.

Graduation committee:

Chairperson:	: Prof. dr. S. Roeser	Section Ethics and Philosophy of Technology
First Supervisor:	: Dr. J. Durán	Section Ethics and Philosophy of Technology
Second Supervisor:	: Dr. L. Rook	Section Economics of Technology and Innovation
External Supervisor:	: D. van Roode	Public Policy Manager at NLdigital



This page is intentionally left blank.

Acknowledgments

Almost seven years ago I arrived in Delft for the start of a new chapter. Throughout the past years, I have seen many people start and finish their master's thesis. Having seen their struggle, I always envisioned writing a thesis as this endless process. Although the process has certainly provided its challenges along the way, since day one, this thesis project has truly been one of the busiest, challenging, and out of comfort zone experiences I have had so far. For the first time, I had to rely on my personal judgment and capabilities for a project this large and long in time span. I have thoroughly enjoyed the entire project and all the opportunities it has provided me with. Along the way, I believe to have developed myself to somewhat of an expert in this niche research field of applied ethics in modern artificial intelligence systems and projects. The fact that I was welcomed for speaking at important events and meetings such as with distinguished professors, captains of the ICT-industry, and members of parliament has given me tremendous confidence in my work and capabilities.

Of course, the honor of this research project does not solely belong to me and recognition is deserved by my well-respected supervisors. First, I would like to thank dr. Juan M. Durán for being such a committed supervisor. Juan has trusted me with the freedom to execute the project in the way I want to but has guided me whenever necessary. Especially the time and commitment he has taken to provide quality and in-depth feedback is admirable. Besides the great atmosphere and collaboration in the project, we have also had some valuable discussions for the rest of my career. Second, I would like to thank my chair Prof. dr. Sabine Roeser, and second supervisor dr. Laurens Rook. From day one, both Sabine and Laurens have shown great trust, provided valuable feedback, and offered to help me whenever necessary. The combination of their expertise, namely Sabine in ethics, and Laurens in research methodology, has provided the necessary perspectives that were needed for this dissertation.

I also want to thank NLdigital, and in particular Dirk van Roode for providing me with whatever means necessary to carry out my work. Dirk has put great effort in arranging meetings, presentations, interviews, and fully trusted me with his professional network. I have felt very welcome at NLdigital since the first day, and I am very thankful for all the opportunities they have provided me with. On top of that, the atmosphere in the office in Breukelen was always great!

Finally, I would like to thank my family, friends, and fellow students. The support of friends and family has been a great encouragement, and their expectations have led me to this ambitious result. The valuable coffee and lunch discussions with fellow students around the thesis project have been of great help in the process. I hope you all enjoy reading this thesis.

Arnoud Nederpel

Delft University of Technology

Executive summary

Background

Artificial intelligence (AI) offers opportunities to solve complex societal and economic challenges. For example, AI-systems could help recognize and cure diseases, reduce traffic incidents, and take over dangerous and repetitive tasks. In the coming decades, AI technologies will only develop further and will have a significant impact on the way many businesses operate. One place where its potential is recognized is The Netherlands. If businesses can continue or accelerate the development of AI-technologies, this could give Dutch businesses a strong competitive advantage, as well as a huge boost for the Dutch economy (e.g., double GDP growth by 2035). However, further progress faces some significant challenges. First, technical limitations such as the need for immense computational power, the lack of well-labeled data sets, and the limitations of the current algorithms hamper its adoption. Another set of fundamental challenges, especially for society, are ethical challenges. Novel challenges for AI arise due to automated decision making, the processing of ever-larger quantities of data, and the complexity of machine learning algorithms. These give rise to ethical challenges related to fairness, explainability, accountability, safety, transparency, and privacy that play a novel role outside of the realm of traditional information systems.

Unfortunately, tools for assessing and guiding the ethical application and development of AI-systems are still in their early stages. Especially, small and medium-sized companies have few resources and guidelines at their disposal. To assist such organizations, NLdigital has created the country's first code of ethics for AI development and application in the ICT-sector. Strikingly, the code was designed without extensive knowledge of the state of AI or knowledge of the ethical challenges and practices in the Dutch ICT-sector. Moreover, only a small group of experts and executives from large organizations were consulted in the creation of the code. The code is considered a living document and the next iteration should be based on relevant data and better represent the practical challenges of small, medium and large corporations.

Purpose and main research question

Due to the lack of such information (i.e., knowledge of the state of AI, ethical challenges, and use of the code of ethics in the sector), it is difficult to establish whether the current code of ethics matches the practical challenges in the industry as well as whether the code is sufficient as an ethical safeguard. It is the purpose of this research to analyze if the code can be improved based on relevant industry data (including from SMEs) and if other important measures for ethical AI governance are missing in the current landscape. This research aims to answer the following research question:

Which ethical challenges are currently addressed insufficiently in the current code of ethics from NLdigital and are fundamental for guiding responsible research and development in the use and application of AI-systems for the Dutch ICT-industry?

Apart from identifying areas for improvement in the code of ethics, this research addresses two knowledge gaps in the literature. First, it contributes to the small body of applied ethics research

on AI in a business setting. Second, it contributes a higher-level descriptive overview of concrete general measures to enforce ethical behavior with AI-systems within companies. Specifically, this research adds a descriptive perspective on ethical governance of AI, and parts can be treated as a case study of the ‘pillars of ethical governance’ framework by Winfield and Jirotko (2018).

Methodology

The methodology that is used to carry out this research is a mixed-method approach, i.e., the *exploratory sequential method*. This approach is the preferred method because it will ensure that the quantitative results have a more contextual basis and are grounded in the experiences of the population (in our case businesses in the Dutch ICT-industry). Specifically, our methodology combines an extensive literature review, semi-structured interview, and an electronic questionnaire. In the literature review, the theoretical context is set, an understanding of the different possible ethical challenges of AI is created, and NLdigital’s code of ethics is analyzed. Subsequently, interviews with different ICT-companies are done to gain practical descriptive insight into the ethical aspects described in literature such as ethical challenges (e.g., fairness, responsibility, safety, etc.) and ethical governance (e.g., code of ethics, whistleblowing channels, ethics training, etc.). Knowledge and experiences from the interviews then inform the quantitative part of this research by functioning as the primary source for the development of the questionnaire. Eventually, the three methods (i.e., literature review and theoretical assessment, interviews, and questionnaire) are triangulated to arrive at the main conclusions of this work and a comprehensive answer to the research question.

Literature review

In the literature review on AI, we explore both the technical domain and theoretical (ethical) domain. First, we adopt a working definition of AI, i.e., machines that are programmed to interpret external data, learn from such data, and use the learnings to successfully carry out a single task. This limits the scope of the technical artifact to narrow AI, even if systems appear to be operating in a much more sophisticated way than that. To describe the theoretical domain, we first provide a review of *consequentialism*, *deontological ethics*, and *virtue ethics* as the main views in normative ethics. However, due to the descriptive nature of this research, moral pluralism was deemed a more useful perspective. By acknowledging that different relevant moral principles and viewpoints exist for ethical challenges of AI, we allow the opportunity to make these conflicting norms, values, and viewpoints visible. However, in the pursuit of a better code of ethics, this research will defend specific *a priori* agreeable moral values. In particular, sufficient literature supports transparency, fairness, explainability, safety and non-maleficence, privacy, responsibility, and accountability as important moral values in AI ethics. Other important concepts that resulted from the literature review and are used throughout the next phases of this research are the necessity of codes of ethics, whistleblowing structures, and pillars of ethical governance.

Following the analysis of NLdigital’s code of ethics, it is concluded that the code is strong with regards to transparency and privacy but is inadequate and lacks practicality with regards to explainability (principle 1) and fairness (principle 3). Moreover, it fails to address safety and non-maleficence issues altogether.

Interview results

A total of six semi-structured interviews have been executed in order to gain an in-depth understanding of the AI and ethics practices of organizations within the ICT-industry. The most pressing ethical challenges for companies in the ICT-industry in the Netherlands that were found during the interviews are fairness and explainability. Challenges with fairness are mainly related to the mitigation and prevention of unwanted biases. On the other side, the key difficulty with explainability was that companies found it hard to explain a complex system or decision and to determine the level of understanding of a target audience that is required. Safety and privacy were also named as important ethical challenges for companies in the sector. Safety was mostly about preventing society from harm and is a difficult challenge because of its unexpectedness and unpredictability. Lastly, although privacy is one of the challenges that companies most frequently have to deal with, the GDPR provides a clear framework for organizations.

In the second part of the interviews, organizations were asked to show the pillars of ethical governance that underlie their organization. It was observed that SMEs have significantly less of such pillars and activities in place than large organizations. The questionnaire further evaluates the extent to which SMEs possess such pillars and internal mechanisms. Finally, feedback on NLdigital's code of ethics was collected. Most frequently, the interviewees mentioned that there is no 'good' principle on fairness included in the code of ethics. Another principle that was seen as inadequate was around the ease of explanation, for example, organizations missed information about the target group and the level of detail that should be included.

Questionnaire results

The primary goal of the questionnaire was to provide a descriptive analysis with regard to the abovementioned concepts. In terms of the general landscape of artificial intelligence in the ICT-sector, the results indicate that 79% of the ICT-businesses active in the Dutch sector use AI for their products, services, or internal processes. In line with the interview results, explainability was considered the most pressing challenge for businesses. Of all the businesses, 56% indicated that explaining predictions, recommendations, or decisions is the most pressing ethical challenge with AI for their business. The results of the questionnaire also demonstrate fairness and privacy among the most pressing challenges in the industry. The questionnaire thus reinforces and validates the interview results that these are challenges that both large businesses and SMEs struggle with.

For dealing with these challenges and to enforce ethical behavior with AI within their business in general, ICT-organizations use different measures or mechanisms. The survey points out that out of the measures discovered during the interviews, a total of five different measures are present within more than half of the businesses, i.e., an open culture to discuss AI ethics, active collaboration with the client, a feedback mechanism for users, strong ethical leadership, and peer-reviewing each other's work. The analysis of pillars of ethical governance for companies in the industry showed that large companies have vastly more ethical pillars on which they built their ethics practices than SMEs. In total, 67% of all SMEs possess zero pillars and none of the SMEs had more than two out of the five pillars of ethical governance in place.

Conclusions

The main research question is answered by triangulating the literature review, interview results, and questionnaire results. The literature review found 1) fairness, 2) responsibility, accountability, and trust, 3) privacy, 4) safety and non-maleficence, 5) transparency and explainability to be the most important ethical aspects. Then, an analysis of NLdigital's code of ethics and a comparison with other codes lead to the identification of three shortcomings, i.e., an unpractical principle of fairness, an unclear principle of explainability, and the absence of a principle for safety. Subsequently, the interviews found that exactly these aspects (i.e., fairness, explainability, and safety) lie at the basis of the most pressing challenges for businesses in the sector. The results of the online questionnaire strongly reinforced these results, especially the importance of explainability and fairness. Lastly, in the interviews businesses indicated that it is precisely these aspects that the code of ethics fails to address adequately. We, therefore, arrive at the conclusion that explainability, fairness, and safety are currently addressed insufficiently in the code of ethics from NLdigital and are fundamental for guiding responsible research and development in the use and application of AI-systems for the Dutch ICT-industry. Based on these findings, I briefly discuss recommendations for a new iteration of the code of ethics and additional governance measures.

Another important finding of this research is the observation that small and medium-sized enterprises are missing concrete measures for dealing with ethical concerns. Considering 1) the large amount of businesses that have started to work with AI-systems, 2) the scalability and relatively large impact of AI-technology, and 3) the absence of concrete pillars, in particular, a code of ethics, a whistleblower mechanism, ethics and RRI training, ethical risk assessments, and transparency about ethical governance, this is cause for concern. Nonetheless, the state of ethical governance within SMEs is not surprising given the early stage these companies are often in and the limited amount of resources they have relative to larger ICT-organizations.

Future research

The relevance of this research is threefold, that is it contributes to the scientific community, industry, and society. Academically, this research adds a practical applied perspective to the body of research on AI ethics, which has previously mainly been discussed on a more abstract level. For the industry, a good ethics code eases daily ethical decision making within companies. Finally, for society, good (self) regulation and ethically responsible AI-practices in businesses will mean it can benefit from the advantages that AI can bring in an ethically responsible manner.

In addition, this work opens new opportunities for future academic research. First, further research could focus on the normative considerations in the development of the next iteration of the code of ethics. Second, a similar descriptive study can be done for different industries such as retail, consumer electronics, automotive, etc. Another possibility for future research is a replication of this work in a few years' time. Finally, this research has recommended the development and provision of additional mechanisms (e.g., training, whistleblowing structures, and ethical risk assessments), especially for small and medium-sized enterprises in the ICT-sector. More research is unquestionably needed to develop these mechanisms and tools to encourage ethical behavior with AI.

Table of Contents

Acknowledgments.....	I
Executive summary.....	II
List of tables.....	IX
List of figures	IX
List of abbreviations and acronyms.....	X
1. Introduction.....	1
1.1 Research background.....	1
1.1.1 The promise of Artificial Intelligence.....	1
1.1.2 The challenge of AI: Ethics	2
1.1.3 A promising tool: NLdigital's code of ethics	3
1.2 Knowledge gap	4
1.3 Problem statement.....	5
1.4 Research questions	5
1.5 Scientific and practical relevance	6
1.6 Research outline	8
2. Research methods.....	9
2.1 Research approach.....	9
2.1.1 Mixed methodology literature review.....	9
2.1.2 Motivation of research methodology for research questions.....	10
2.1.3 Research flow	13
2.2 Research design.....	14
2.2.1 Interviews	14
2.2.2 Questionnaire	16
2.3 Conclusion.....	19
3. Literature review of artificial intelligence and ethics.....	21
3.1 Technical domain.....	21
3.1.1 Artificial intelligence	21
3.1.2 Machine learning	24
3.1.3 Design and implementation of Artificial intelligence.....	26
3.2 Theoretical domain	28
3.2.1 Ethics theory	29

3.2.2	Ethics of artificial intelligence	30
3.2.3	AI Ethics governance	38
3.2.4	Codes of Ethics.....	40
3.3	Ethical guideline delineation.....	45
3.3.1	The landscape of AI principles and guidelines	45
3.3.2	Analysis of NLdigital's ethical code	46
3.3.3	European Union ethical guideline.....	50
3.3.4	Various enterprise ethics principles.....	51
3.4	Conclusion.....	53
4.	Interview results.....	56
4.1	The landscape of Artificial Intelligence	56
4.2	Important ethical challenges with AI in the industry	57
4.2.1	Most pressing ethical challenges	57
4.2.2	Other ethical challenges	60
4.2.3	Interpretation of colleagues.....	61
4.3	Mechanisms for dealing with ethical challenges and increasing ethical behavior	62
4.3.1	Mechanisms for dealing with ethical challenges.....	63
4.3.2	Internal mechanisms for increasing ethical behavior	65
4.3.3	Pillars of ethical governance	66
4.4	Feedback on the current code of ethics	68
4.4.1	Usefulness of principles	68
4.4.2	Missing principles	68
4.4.3	Formulation of the principles	69
4.4.4	Adoption and impact.....	70
4.5	Conclusion.....	71
5.	Questionnaire results	73
5.1	Exploratory sample analysis	73
5.1.1	Data cleaning.....	73
5.1.2	Exploratory descriptive statistics	73
5.2	The landscape of artificial intelligence.....	75
5.2.1	Businesses without AI-systems for products, services, or internal processes	76
5.2.2	Businesses using AI-systems for products, services, or internal processes	76
5.3	Pressing ethical challenges.....	78

5.4	Ethical governance	80
5.4.1	Internal mechanisms for increasing ethical behavior	81
5.4.2	Pillars of ethical governance	82
5.5	Code of ethics.....	85
5.5.1	Adoption of the code	85
5.5.2	Adherence to the code.....	85
5.5.3	Feedback statements.....	86
5.6	Conclusion.....	87
6.	Conclusions	90
6.1	Main findings.....	90
6.1.1	Towards an improved code of ethics: answer to the research question.....	90
6.1.2	Other notable findings.....	94
6.2	Limitations of this research	97
6.3	Further discussion.....	98
6.3.1	Normative ethical reflection on the main findings	98
6.3.2	Discussion on ethical governance of AI within organizations.....	99
6.4	Recommendations	100
6.4.1	The code of ethics	100
6.4.2	Additional mechanisms	100
6.5	Future research	102
6.6	Link of this work with the Management of Technology program	103
	References	104
A.	Interview protocol.....	109
B.	Interview analysis Atlas.ti	112
C.	Questionnaire.....	117
D.	Data treatment.....	123
D.1.	Renaming variables	123
D.2.	Removing and adding variables	125
D.3.	Cleaning of test responses, incomplete responses, and outliers	126

List of tables

Table 1: Research method per sub-research question	10
Table 2: Interview respondents	15
Table 3: Margin of error on different confidence intervals of the electronic questionnaire	19
Table 4: Machine learning methods	25
Table 5: Ethical challenges of artificial intelligence emergent in different studies	31
Table 6: Overview of different responsibility concepts	34
Table 7: Types of passive responsibility	34
Table 8: Analysis of 84 ethical principles and guidelines, adapted from: (Jobin et al., 2019)	46
Table 9: NLdigital's code of ethics and ethical building blocks.....	48
Table 10: Ethical requirements for trustworthy AI, adapted from: (HLEG AI, 2019)	50
Table 11: Enterprise ethics principles of IBM, Microsoft, and Google	51
Table 12: Conceptual overview of ethical challenges with artificial intelligence	54
Table 13: Most pressing ethical challenges resulting from the interviews	58
Table 14: Mechanisms and activities for increasing ethical behavior in the ICT-sector.....	65
Table 15: Presence of pillars for good ethical governance in the interviewed companies.....	66
Table 16: Most pressing ethical challenges with AI for businesses in the Dutch ICT-industry....	80
Table 17: Frequency of internal mechanisms to encourage ethical behavior.....	81
Table 18: Renamed variables and question number	123
Table 19: Removed variables.....	126
Table 20: Added variables	126

List of figures

Figure 1: Research flow diagram	13
Figure 2: Artificial intelligence subdivision, adapted from: (N. Singh, 2016).....	23
Figure 3: Layered model for AI governance, adapted from: (Gasser & Almeida, 2017)	39
Figure 4: Distributions of respondent's organization by employees and company size.....	74
Figure 5: Company headquarters by size.....	74
Figure 6: Company footprint by size.....	75
Figure 7: Usage of AI within ICT-businesses in the Netherlands	76
Figure 8: Stage of development of AI-systems in the Dutch ICT-industry	77
Figure 9: Autonomy of development of AI-systems	77
Figure 10: Most pressing ethical challenges with AI for businesses in the Dutch ICT-industry by size from the questionnaire	79
Figure 11: Internal ethical mechanisms by company size.....	82
Figure 12: Number of pillars of ethical governance in organizations by size	83
Figure 13: Presence of individual (parts) of pillars of ethical governance	84
Figure 14: Familiarity of businesses with NLdigital's code of ethics.....	85
Figure 15: Extent to which organizations follow the principles of the code of ethics	86
Figure 16: Feedback on code of ethics statements measured on a Likert scale	87

List of abbreviations and acronyms

AI	Artificial Intelligence
AGI	Artificial General Intelligence
CEO	Chief Executive Officer
CSR	Corporate Social Responsibility
EU	European Union
GDPR	General Data Protection Regulation
IEEE	Institute of Electrical and Electronics Engineers
ICT	Information and Communication Technology
ML	Machine Learning
MLaaS	Machine Learning as a Service
RRI	Responsible Research Innovation
SME	Small and Medium-sized Enterprise
SRQ	Sub Research Question

1. Introduction

This chapter introduces the research problem. First, the background and context of this research are set in section 1.1. This section is followed by the identification of the knowledge gap (section 1.2). Then, section 1.3 offers the research problem statement that is translated into the main research question and six sub-research questions in section 1.4. The scientific relevance of this research and practical relevance for the ICT-industry are delineated in section 1.5. Finally, section 1.6 provides a brief introduction to the research methodology and lays out the overall outline of this research.

1.1 Research background

The sections below discuss the promises of artificial intelligence (section 1.1.1) and the challenges the technology and its development face both technically and societally (section 1.1.2). Finally, in section 1.1.3 NLdigital's code of ethics is presented as a tool that can help overcome some of these challenges.

1.1.1 The promise of Artificial Intelligence

Present-day, artificial intelligence is one of the most popular topics in business and academic research. The scientific field of AI has seen more than a nine-fold growth in the annual publishing rate of academic papers in the last two decades (Shoham et al., 2018). This is because artificial intelligence (henceforth AI) offers opportunities to solve complex societal and economic challenges all around the world. Google CEO Sundar Pichai even compared its potential impact with that of humanity's most important inventions such as fire and electricity (Goode, 2018). Pichai, who is widely recognized as one of the experts in the field, believes AI technologies could be used to cure cancer and solve important climate change challenges in the future. More recently, new AI techniques such as machine learning and its subset deep learning have made it possible to let algorithms make sense of massive quantities of data without dictating a large set of rules and constraints. This has enabled free open-source AI software to spot anomalies on x-rays (Wilson, 2019), Netflix to recommend and create content seamlessly to the likes of their customer (Plummer, 2017), and most recently it has enabled a new system called 'Aristo' to pass an 8th-grade science test (Metz, 2019). In the coming decades, AI technologies will only develop further and will have a significant impact on the way many businesses operate.

Artificial intelligence offers a multitude of benefits for society. For example, self-driving cars can vastly reduce the number of accidents and casualties in traffic, save time, and open the possibility for completely new mobility concepts. Robots can take over numerous tedious and dangerous tasks from humans, with virtually no error. Yao, Zhou, and Jia (2018, p. 31) also observed that artificial intelligence systems could bring rapid innovation and progress to microfinance, social justice, and medical diagnosis, among many other branches of knowledge. The Netherlands has historically been a country at the forefront of many technological developments. Likewise, businesses in the Netherlands have been among the first to adopt AI technologies in their business practices. Incumbents such as TomTom and booking.com are pioneers in their field, and the Netherlands is now home to over 200 innovative AI start-ups. By positioning themselves strategically, the Netherlands could reap the societal benefits of AI (AINED, 2018).

Economically, AI technologies can have a major impact as well. If the Netherlands can continue or accelerate the development of its position as one of the global leaders in AI, this would mean a huge boost for the Dutch economy. A 2017 research from PWC concluded that “AI could contribute up to \$15.7 trillion to the global economy in 2030” (Rao & Verweij, 2017, p. 2). Another research by Accenture shows that the Netherlands can grow its labor productivity by 27% more compared to the baseline that is expected by 2035, provided that it was to absorb more AI into the economy. As a result, the Dutch economy could realize an annual growth rate of 3.2% in this scenario, compared to the otherwise projected 1.6% (Purdy & Daugherty, 2016).

1.1.2 The challenge of AI: Ethics

Despite the fact that artificial intelligence offers many opportunities for the advancement of businesses and society, further progress faces some significant challenges. First, there are technical challenges that inhibit the full potential of AI. Vast amounts of computational power are needed to run the state-of-the-art machine learning and deep learning algorithms (Adixon, 2019). However, current systems with massive processing power such as cloud computing and parallel processing are reaching their current limits. Furthermore, well-labeled data sets that are large enough for creating accurate AI applications are scarce in most businesses and industries. Innovation in both domains is needed for the further development of AI.

However, the most pressing set of challenges arises in the societal field. Over the last decade, many important books on artificial intelligence have included a chapter dedicated to the ethical aspects (Bostrom & Yudkowsky, 2014, pp. 316-334; Müller, 2018, pp. 231-315; Russell & Norvig, 2016, pp. 1034-1040; Smith & Browne, 2019, pp. 191-211; Yao et al., 2018). However, widespread ethical usage of artificial intelligence seems to currently be far from reality. Last year, for example, Amazon had to shut down its AI-powered recruiting tool because of latent bias. The machine learning algorithm that they used was trained with resumes from predominantly men and as a result taught itself that men were better candidates for Amazon’s open positions than women (Dastin, 2018). Biased AI systems that discriminate based on gender, ethnicity, color, or income are a serious ethical issue that needs to be tackled for AI’s sustainable development. Likewise, numerous other ethical challenges exist for artificial intelligence-based systems. Novel challenges for artificial intelligence arise due to automated decision making, the processing of ever-larger quantities of data, and the complexity of machine learning algorithms. These causes ethical challenges to play a new role outside of the realm of traditional information systems. In particular, fairness, explainability, accountability, transparency, privacy and safety are hypothesized to be pressing ethical challenges for businesses in the industry.

Unfortunately, tools for assessing and guiding ‘ethical behavior’ in artificial intelligence are still in their early stages. Existing overarching tools such as the guideline of technical professional association IEEE and the European Union guideline for trustworthy AI (HLEG AI, 2019; IEEE, 2016), are not tailored to the specific challenges of industries. Thus, companies like Google, IBM, and Microsoft have formulated their own ethics principles for AI, and have developed tools for testing their systems on biases (Google, 2019; IBM, 2018; Microsoft, 2019). Nevertheless, small and medium-sized companies have few tools and guidelines at their disposal that can help them assess their AI-systems, responsibilities, and ease daily decision making.

1.1.3 A promising tool: NLdigital's code of ethics

The General Data Protection Regulation (GDPR) was implemented by the European Union in 2018 to protect the data privacy of European citizens (EU, 2016). Although the GDPR provides some protection for data privacy in information systems, specific laws for artificial intelligence do not exist yet in the Netherlands. Experimental regulation is now in the process of special use cases such as autonomous driving in the pursuit of good future regulation (TNO, 2019). It will take significant time before regulation for other ethical challenges such as responsibility, fairness, and safety for artificial intelligence will pass in the Netherlands. Other tools and mechanisms are needed for companies in the meantime. Yao et al. (2018, pp. 43-44) observe that guidelines and codes of conduct can be an important predecessor of good policy. Myriad researchers in the last two decades have tried to establish the effectiveness of codes of conduct (Adam & Rachman-Moore, 2004; Erwin, 2011; Paine, 2004). A large body of the literature agrees that if ethical codes are designed and implemented adequately, the code may have a significant impact on the ethical behavior of people in an organization (Adam & Rachman-Moore, 2004; Erwin, 2011). Thereby it becomes an essential factor in the success of business corporations (Paine, 2004).

To guide ICT businesses in the Netherlands through the ethical challenges, NLdigital, the trade association for the ICT-sector in the Netherlands, together with some of its larger associated companies, has created the country's first ethical code of conduct for AI application (NLdigital, 2019). The code consists of eight principles and aims to promote the ethical application of AI in organizations. The code was a result of cooperation over the course of around a year with several members of the association. After a series of iterations, ultimately, the 'Ethical Code for Artificial Intelligence' was accepted by the board of NLdigital. Peter Zijlema (GM of IBM Netherlands) handed the code over that was printed on a mirror on the 21st of March to State Secretary for Economic Affairs and Climate Policy Mona Keijzer. The code of ethics is a non-binding code and was designed specifically but not solely for its members. The goal was to make the guideline broadly applicable. For this reason, NLdigital has deliberately used a broad definition of artificial intelligence in the creation of the code, i.e. everything that touches self-learning or self-reasoning systems (van der Beek, 2019). According to NLdigital, the guideline is a living document that may be reviewed "as the need arises" (NLdigital, 2019).

Nonetheless, the ethics code was designed without extensive knowledge of the state of AI and ethical practices in the Dutch ICT sector. This is because, even though the literature on AI in businesses and AI ethics from a philosophical perspective is plentiful (Yao et al., 2018), industry-wide data on the ethical application of AI is scarce, especially for the ICT industry in the Netherlands. On top of the fact that the code was designed without much practical data from the industry, only a small group of executives mostly from large multinational corporations was consulted in the creation of the code. Although this means the code reflects the experiences of high-level management, herein also lies a caveat. The ethical code reflects the perception of the experts of a select group of companies and thereby neglects the perspective of the employees that make most of the practical daily decisions. Furthermore, the input for the code came from a relatively small sample of which predominantly large multinational corporations, posing limitations on the generalizability of the code. This is a crucial aspect because the code was designed for the entire industry. A next iteration should be more generalizable and better represent the practical challenges of small, medium and large corporations. Our assumption is that especially

with regards to important ethical challenges such as explainability, fairness, safety, responsibility, the code can be improved. Finally, it is unknown if an improved code by itself is sufficient as a measure to encourage ethical behavior or deal with ethical challenges. In response to these challenges, this study proposes a comprehensive analysis of the state of AI and the ethical challenges with AI in the Dutch ICT-sector so that the code of ethics can be revisited based on data from the entire industry. This research also provides an exploratory study on the need for additional measures, besides a code of ethics.

1.2 Knowledge gap

Ethical challenges are the key (societal) issues that need to be dealt with to unleash the full potential of artificial intelligence for society (Lufkin, 2017). Companies must find a balance between ethical questions such as privacy, auditability, safety, transparency, and fairness. Fundamental to all those challenges are the questions of explainability and responsibility. Enterprises, trade associations, and policymakers need to think carefully about how to deal with these fundamental challenges.

For the Dutch ICT industry, AI is becoming an increasingly more important topic and business asset. This industry is an important one for the Netherlands because it employs around 370.000 professionals in over 50.000 companies (ABN-AMRO, 2019). To guide ICT companies in the Netherlands in the uncertain landscape of AI, a code of ethics was constructed. However, the input for the code came from a relatively small sample of experts and mainly originating from large multinationals, posing limitations on the generalizability of the code. This is a crucial aspect because the code was designed for the entire industry. Furthermore, little statistical information on ethical challenges, safeguards and the adoption of the ethics code in the industry is available, making it hard to measure the relevance and impact of the code altogether. This is unfortunate because an effective code of ethics can benefit the industry and country. This approach was sufficient for the first version of the code, but now NLdigital wants to verify their code of ethics based on further analysis. They have explicitly formulated the need for a descriptive analysis that maps the ethical challenges of the businesses emerging in the sector. This data will enable measurement of the 'performance' (adoption, feedback) of the code, and makes it possible to revisit the code based on data from the entire industry. Apart from identifying areas for improvement in the code of ethics, this research addresses the following gaps in the literature:

- **Applied ethics of AI research:** the body of applied ethics research on artificial intelligence in business settings is relatively small. Especially, descriptive research with respect to ethical challenges and their degree of importance from the perspective of business is an angle that remains relatively unexplored. In particular, such research for the Dutch ICT-industry is extremely scarce. This research aims to fill these gaps by doing a descriptive study on ethical challenges within businesses in the Dutch ICT-industry.
- **Research on concrete measures and ethical governance of artificial intelligence:** the body of research on concrete general measures to enforce ethical behavior with artificial intelligence systems within companies is either reduced or it has focused on specific topics such as codes of ethics and education. Researches that provide a higher-level descriptive overview of what businesses actually use instead of what they should use is scarce. Furthermore, even though research on ethical governance exists, such as

the research of governance pillars of Winfield and Jirotko (2018), little research has been conducted to test the degree to which ethical governance of AI exists in practice. Specifically, a case study or descriptive perspective of the 'pillars of ethical governance' framework is missing for the ICT-industry. This research aims to fill this gap by discovering which measures and governance pillars that are described as important in contemporary academic literature (Boddington, 2017; Winfield & Jirotko, 2018), are also present in practice.

1.3 Problem statement

In light of the research background (section 1.1) and the knowledge gap (section 1.2), the research problem statement is formulated as the following:

Bringing more artificial intelligence into practice in an ethically thoughtful way will offer the Netherlands many societal and economic benefits. The code of ethics from NLdigital guides organizations within the ICT industry through the ethical challenges that such technologies bring. The code is a living document that requires continuous improvement. The first version of the code was made without extensive knowledge of the development and application of AI-systems in the industry and its inherent ethical challenges. Rather, it was a collaboration of a handful of experts and large corporations. A next iteration should be more generalizable and better represent the practical challenges of small and medium-sized corporations. However, the next iteration cannot be made because little is known about the ethical challenges of these companies in the industry and the areas for improvement in the current code of ethics. Moreover, it is unknown if the code is sufficient as a measure to encourage ethical behavior because the code's value and relevance to practice have not been researched yet. In response to these problems, this study proposes a *descriptive analysis* of the state of AI and the ethical challenges with AI in the Dutch ICT sector. This research will also map out the ethical safeguards, measures, and governance that currently protect companies from ethically problematic decisions and identify opportunities for improvement. Ultimately, these different angles of knowledge are brought together to find out how the code of ethics can be improved.

1.4 Research questions

Evidently, by the articulation of NLdigital, there is a need for more descriptive knowledge on AI-practices in the industry to develop the next iteration of the code of ethics. The sections above support the statement that such information specifically about the Dutch ICT-industry is scarce. Due to the lack of such information, it is hard to establish if the current code of ethics matches the practical challenges in the industry and if the code is sufficient as a safeguard. This knowledge gap leads this research to adopt the following research question:

- **Which ethical challenges are currently missing in the current code of ethics from NLdigital and are fundamental for guiding responsible research and development in the use and application of AI-systems for the Dutch ICT-industry?**

To give a comprehensive answer to this research question the following sub-research questions will be answered:

1. What are the chief ethical aspects in the research, development, and application of AI-systems for businesses in the Netherlands that are identified in the specialized literature?
2. What ethical aspects in NLdigital's code of ethics are missing that can be identified by comparing it to other codes of ethics?
3. How do companies in the Dutch ICT-industry perceive the ethical challenges of AI?
4. Which ethical challenges are most pressing to Dutch ICT-businesses in the development and application of AI-systems?
5. What concrete measures do Dutch ICT-businesses currently use to deal with the ethical aspects of developing and applying AI-systems?
6. What are the common concerns that are uttered by the Dutch ICT-community about the current code of ethics offered by NLdigital?
7. What concrete measures are missing that would assist small and medium-sized companies in dealing with ethical challenges, and how does a code of ethics play a role here?

1.5 Scientific and practical relevance

The relevance of this research is threefold, that is, to contribute to the scientific community by providing a descriptive applied perspective of AI-ethics and governance, to the industry by providing areas for improvement in NLdigital's code of ethics, and to policymakers by creating knowledge and data that are necessary for informed discussion/ and decision making in the design of future strategy and regulation. The following sub-sections lay-out the concrete contributions of this work to each one of the aforementioned fields.

1.5.1.1 *Academic contribution*

Academically, this research adds an applied perspective to the body of research on AI ethics. Much research can be found that discusses the importance of ethics in AI, ways in which we can deal with ethical challenges with AI, and the value of a code of ethics (for AI). However, few descriptive studies have been done to map out precisely these aspects. Especially, descriptive research with respect to ethical challenges and their degree of importance from the perspective of business is an angle that remains relatively unexplored. This study contributes with such a descriptive study of the ICT-sector, which is the leading industry with regards to the development and implementation of AI-systems in the Netherlands. On top of the research on ethical challenges in the industry, this work provides descriptive insight into the way in which AI is currently governed in organizations and what measures businesses use to encourage ethical behavior with artificial intelligence. This is a topic that is currently under discussion on an abstract level rather than from an applied perspective. Researches that provide a higher-level descriptive overview is scarce. This work fills the abovementioned knowledge gaps by mapping out the tools and mechanisms that are used in the industry (e.g., whistleblowing channels, training, ethics communities, etc.) and what small and medium-sized enterprises can learn from the way multinationals govern AI-ethics. Furthermore, this research provides a case study of the 'pillars of ethical governance' framework, namely for the ICT-industry (Winfield & Jirotko, 2018).

The results of this study can then be used for other AI-related research in the sector such as policymaking, strategy development, development of additional tools and mechanism to enable, encourage, and enforce ethical behavior, and different ways to administer the ethical practice of

AI within enterprises or networks of enterprises. Lastly, because the research is descriptive in nature and measures a wide range of ethics practices at one point in time, the research can be replicated to track the changing landscape of AI in the Netherlands. This could, in turn, determine the effectiveness of different policies and open new areas for research.

1.5.1.2 Industry contribution

The results of this research will identify opportunities for improvement and will be used to develop the next iteration of the code of ethics for the industry. This means this work directly helps towards a better and more widely adopted ethical code that will protect companies from negative repercussions such as liabilities and reputational damage. For the industry, a good ethics code eases daily ethical decision making within companies. However, the intention of this thesis is not to provide the normative dimension of such a next generation of code of ethics. Instead, we will lay out the basis for such a second iteration and provide some first suggestions. Second, this research identifies additional mechanisms that are necessary to enable, encourage, and enforce ethical behavior with artificial intelligence systems in the ICT-industry. Specifically, for small and medium-sized enterprises an important contribution is the consciousness that such mechanisms exist and are necessary, as well as the recommendations this work provides for enabling additional mechanisms by overarching organizations such as NLdigital.

Lastly, the results of the research allow organizations to assess where they stand in the ecosystem, e.g. with regards to their partners and competitors. This work allows them to grasp where they are in terms of AI development against the industry averages and what their partners and competitors do in terms of ethical governance. Finally, it encourages businesses to have discussion and collaboration around ethics in AI with other actors, because this research shows that their ethical challenges (fairness, explainability, and safety) are shared among the rest of the industry.

1.5.1.3 Societal and political contribution

Regulation for artificial intelligence will gain increasing importance for policymakers in the coming years (AINED, 2018). It is the position of our research, that it is vital that such policy is 'right' the first time so that the sustained development of AI in the Netherlands is neither slowed down nor hampered. This means that policymakers should be well informed from a wide range of perspectives and levels. Unfortunately, the perspective of the industry is currently underrepresented in the relevant literature. This research adds such perspective in terms of businesses in the Dutch ICT-sector. Dutch (or European) policymakers can use the knowledge, data, and recommendations resulting from this research in the development of AI-regulation or a national ethical AI-guideline. For example, the identification of missing (but important) governance measures within businesses, should be concrete points of discussion in the design of future strategy and regulation.

For society, good regulation and ethically responsible AI-practices in businesses will mean it can benefit from the advantages that AI can bring in an ethically responsible manner. This signifies profiting from the possibilities that AI opens for our society without being prone to harm, privacy intrusion, unexplainable black-box decision making, unaccountable systems, and unfair distributions.

1.6 Research outline

NLdigital has constructed a code of ethics to guide Dutch ICT businesses in the uncertain landscape of artificial intelligence. Although the code was developed in dialogue with some of the trade association's members and advised by the digital ethical council of the association, the code is built on experiences of just a few companies and experts rather than rigorous analysis. The result of the methodology used is that the majority of the ICT community in the Dutch industry fails to delegate their practices to the code of ethics. Additionally, there is a mismatch between the ethical challenges that the code addresses and the challenges that companies in reality struggle with. Both obstruct the adoption of the code by the ICT community. This approach was sufficient for the first version of the code, but now NLdigital would like to verify or edit the code based on further analysis. Unfortunately, a more rigorous analysis that analyzes the code of ethics and maps the ethical challenges of the businesses emerging in the sector is presently not available to the organization. To address this knowledge gap, this thesis provides a descriptive research of AI-ethics in the Dutch ICT sector based on statistical analysis on interviews and questionnaires.

The structure of this work is as follows; in chapter 2, it is motivated why a mixed-method study approach is the most suitable approach for this research. This chapter also provides a more detailed discussion of the design of this study. In short, this research will consist of two main phases, i.e. a qualitative phase and a quantitative phase. First, the qualitative phase consists of an extensive literature and theoretical review on artificial intelligence and ethics (chapter 3) in which we motivate the importance of ethical aspects such as explainability, fairness, safety, transparency, privacy, responsibility, and ethical governance. Furthermore, this chapter also provides the general ethical theoretical framework for this research. Then, NLdigital's code of ethics is analyzed to identify that explainability, fairness, and safety are not addressed in the code in a meaningful way. Following, in chapter 4, the results of the set of semi-structured interviews are discussed to understand the position of the industry towards the different ethical aspects and to understand their ethical challenges and governance. The information gathered throughout the first phase provides the input for the questionnaire that is carried out in the second phase. The questionnaire, of which the results are discussed in chapter 5, reinforces the results from the interviews. Both methodologies indicate that explainability and fairness are the two most pressing challenges in the industry, followed by safety and privacy. Finally, in chapter 6 we triangulate the different findings to show that fairness, explainability, and safety are indeed addressed insufficiently in the code of ethics, but are fundamental for guiding responsible research and development in the use and application of AI-systems for the Dutch ICT-industry. Although this research is descriptive, we will nevertheless give some normative suggestions for the next generation of code of ethics. Hence, the conclusions of this work are followed by a brief ethical reflection on the main findings, recommendations for a better aligned and more impactful ethics code and ethical governance landscape, and possibilities for future research.

2. Research methods

This chapter lays out the research approach and methodology that are used throughout the study. Section 2.1 elaborates on the overall research strategy for this thesis and includes a visualization of the research flow. Section 2.2 describes the individual methods of the research approach and lays out how these are executed.

2.1 Research approach

This section will describe the scientific approach adopted for this research. Section 2.1.1 and 2.1.2 elaborate on the literature review of possible research methods as well as opts for the mixed-method approach as the most suitable strategy for our purposes and research questions. Section 2.1.3 visualizes the resulting overall flow of this study.

2.1.1 Mixed methodology literature review

This research will make use of the mixed-method approach. Mixed method research is the dominant name for research that integrates both a qualitative and quantitative analysis to derive its conclusions (Creswell, 2013). For this study, the mixed-method approach is a valuable method because it will ensure that the quantitative results have a more contextual basis and are grounded in the experiences of the population (in our case businesses in the Dutch ICT-industry), which is specifically valuable in a normative field such as ethics (Wisdom & Creswell, 2013). The qualitative approach is required to answer the first three sub-questions of this research, namely:

- What are the chief ethical aspects in the research, development, and application of AI-systems for businesses in the Netherlands that are identified in the specialized literature?
- What ethical aspects in NLdigital's code of ethics are missing that can be identified by comparing it to other codes of ethics?
- How do companies in the Dutch ICT-industry perceive the ethical challenges of AI?

In doing this, the theoretical context is set, an understanding of the different possible ethical challenges of AI is created, and the way in which the industry thinks about solving those challenges is understood. This knowledge will inform the quantitative part of this research by functioning as the primary source for the development of the questionnaire that will be sent out to businesses in the Dutch ICT-industry. They will thus form the foundation for the construction of the survey structure and questions that we will be using in the rest of this research.

One of the leading articles in mixed-method research is that of Johnson and Onwuegbuzie (2004) which makes a comparison between qualitative, quantitative and mixed-method research. They argue that mixed-method research takes the strongest aspects of both traditional research paradigms. They also observe that there is not one dominant research design in mixed-method research, which is in strong disparity with the mono qualitative or quantitative paradigm. Rather, in mixed-method research, the researcher is free to use both approaches as he seems fit to effectively answer the research question at hand (Venkatesh, Brown, & Sullivan, 2016). Two important decisions, however, should be made before the start of the research, i.e. "(a) whether one wants to operate largely within one dominant paradigm or not, and (b) whether one wants to conduct the phases concurrently or sequentially" (Johnson & Onwuegbuzie, 2004, p. 20).

Although Venkatesh et al. (2016) and Johnson and Onwuegbuzie (2004) formulations of mixed methods bring up important elements, this research profits best from Creswell's definitions of the exploratory sequential method where data collection occurs in two phases, i.e., a qualitative and a quantitative phase (Creswell, 2013, p. 220). In this study, the qualitative phase will be conducted to inform the quantitative phase. To this end, I will first conduct a literature review on the field of normative ethics, AI ethics, and code of ethics theory to set the theoretical context, motivate important principles in AI ethics, and create a list of ethical challenges with AI. By doing this the core concepts for this research are expounded and clear. Then, semi-structured and open-ended exploratory interviews are used to further conceptualize the ethical challenges that businesses could be experiencing with AI by bringing in a more practical and applied perspective. Interviews will be conducted with the purpose of allowing a suitable interpretation of how the Dutch ICT industry perceives the different ethical challenges discussed in the previous theoretical context. Semi-structured interviews are the preferred interview method because of its strong characteristic to gain insight from the perspective of the interviewee, and because of the flexibility, it allows to explore different domains and underlying concepts. These concepts together with the important ethical challenges that are identified in theory and literature will form the 'building blocks' as we call them in this research for the conceptual framework for the quantitative analysis.

The deliverables from the previous analysis are an overview of how the industry works with AI, a ranking of importance of the different ethical challenges, an overview of the concrete measures the industry uses to govern ethical behavior with AI, and statistical data about the adoption of the code of ethics. A questionnaire is chosen as the main instrument to collect quantitative data because it is a quick method that is easily scalable to a larger sample. This is desirable because of the limited time of the research. NLdigital also has its own platform in place for questionnaires to reach all its members. The approach is thus a sequential one where the dominant research method is the quantitative analysis. Survey instrument development will function as the link between the qualitative phase and the quantitative phase (Creswell, 2013, pp. 225-227). The quantitative phase will ensure the generalizability of the research. For the practical design of the research, the interview protocol (section 2.2.1) and survey design (section 2.2.2) will follow the guidelines of Sekaran and Bougie (2016). The next section summarizes the choice for the specific research methods in the mixed-method design of this study.

2.1.2 Motivation of research methodology for research questions

Each sub-research question is answered using methodologies that are most useful for the nature of that question. Most often this is by combined insights from the literature review, interviews, and questionnaire. Table 1 displays which methods are used for each of the sub-questions.

Table 1: Research method per sub-research question

No.	Section	Sub-research question	Research method
1.	3.2.4	What are the chief ethical aspects in the research, development, and application of AI-systems for businesses in the Netherlands that are identified in the specialized literature?	Literature review

2.	3.3.3	What ethical aspects in NLdigital's code of ethics are missing that can be identified by comparing it to other codes of ethics?	
3.	4.2.3	How do companies in the industry perceive the ethical challenges of AI?	Semi-structured interviews
4.	Part 1: 4.2.1 Part 2: 5.3	Which ethical challenges are most pressing to Dutch ICT-businesses in the development and application of AI-systems?	Semi-structured interviews, questionnaire
5.	Part 1: 4.3.3 Part 2: 5.4.2	What concrete measures do Dutch ICT-businesses currently use to deal with the ethical aspects of developing and applying AI-systems?	
6.	4.4.3	What are the common concerns that are uttered by the Dutch ICT-community about the current code of ethics offered by NLdigital?	
7.	6.1.2	What concrete measures are missing that would assist small and medium-sized companies in dealing with ethical challenges, and how does a code of ethics play a role here?	Semi-structured interviews, questionnaire, literature review

Sub-research questions one and two are first answered using the literature review on AI-ethics and AI codes of ethics. Abundant literature relating to the first sub-question exists. Prominent academic works on AI-ethics and codes of ethics are reviewed to identify the different ethical challenges with AI that are named in the literature. In order to answer sub-question 2, this work will execute an in-depth analysis of the elements that make up NLdigital's code of ethics. Then, the elements and principles that make up other prominent AI codes of ethics (e.g., EU guideline for trustworthy AI) are delineated and finally contrasted with NLdigital's guideline to identify areas for improvement.

The third sub-question is answered following the set of semi-structured interviews. The subtle differences in judgment and perception of the ethical challenges and the way in which the industry deals with them can be better understood using interviews. The interviews will allow the analysis to be focused on the practical challenges and provide an industry-specific perspective.

The fourth, fifth, and sixth sub-questions are answered by a combination of methods, i.e., the semi-structured interviews and the questionnaire. First, sub-question four is partially answered with the results of the interviews. The semi-structured interviews allow for detailed information with regard to the most pressing challenges for businesses. Follow up questions will be asked to thoroughly understand the root cause of these challenges. The challenges that have come up during the interviews are then tested for a wider sample using the questionnaire. Together, the qualitative context and motivation from the interviews and the wider generalization that the questionnaire provides will determine the most pressing ethical challenges in the development and application of AI-systems in the ICT-industry. Second, sub-question five is answered using a

similar methodology as the previous one. Only in this case, ICT-businesses are asked about concrete measures they have in place to deal with the ethical challenges. The interviews partially answer the sub-question by mapping out the different measures used in the industry and providing a comprehensive understanding of what such measures look like and why they are used, and the questionnaire again makes it possible to generalize these results to a wider sample. Third, sub-question six is mainly answered using exploratory semi-structured interviews. Respondents will be given time to study NLdigital's code of ethics and are subsequently asked to provide feedback on the principles in the code and identify areas for improvement such as missing principles. The questionnaire will contribute to this answer by determining the adoption of the code and the willingness to use it if the code is improved with the suggestions given by organizations during the interviews.

Next, sub-research question seven is answered using a combination of the literature review, interviews, and questionnaire. This integration is necessary so that the quantitative results are theoretically and contextually grounded. The theoretical support for important measures for dealing with ethical challenges comes from literature, whereas the contextual motivation follows from the experiences shared by ICT-business during the interviews. Finally, the quantitative data of the actual measures that companies currently have in place is based on the results of the questionnaire. Triangulating these three methods will provide the answer to what concrete measures are missing that would assist small and medium-sized companies in dealing with ethical challenges, and how a code of ethics could play a role here. Finally, the main research question: 'Which ethical challenges are currently missing in the current code of ethics from NLdigital and are fundamental for guiding responsible research and development in the use and application of AI-systems for the Dutch ICT-industry?' is answered by triangulating the literature review and code of ethics analysis, interviews, and questionnaire. The literature review motivates which ethical aspects are important and why they are important. The analysis of NLdigital's code of ethics will identify principles that are not included or not sufficiently addressed in the code but were deemed important in prominent literature. Lastly, the interviews and questionnaire are used to demonstrate that organizations in the Dutch ICT-industry do in fact experience those challenges as pressing and indicate to have insufficient measures to deal with them.

2.1.3 Research flow

The research flow diagram in Figure 1 systematically shows the structure of the research. The flow of the diagram follows the exploratory sequential mixed methodology. The sequential execution of a qualitative phase, quantitative phase, and consequent triangulation of these approaches helps to answer the seven sub research questions and the main research question.

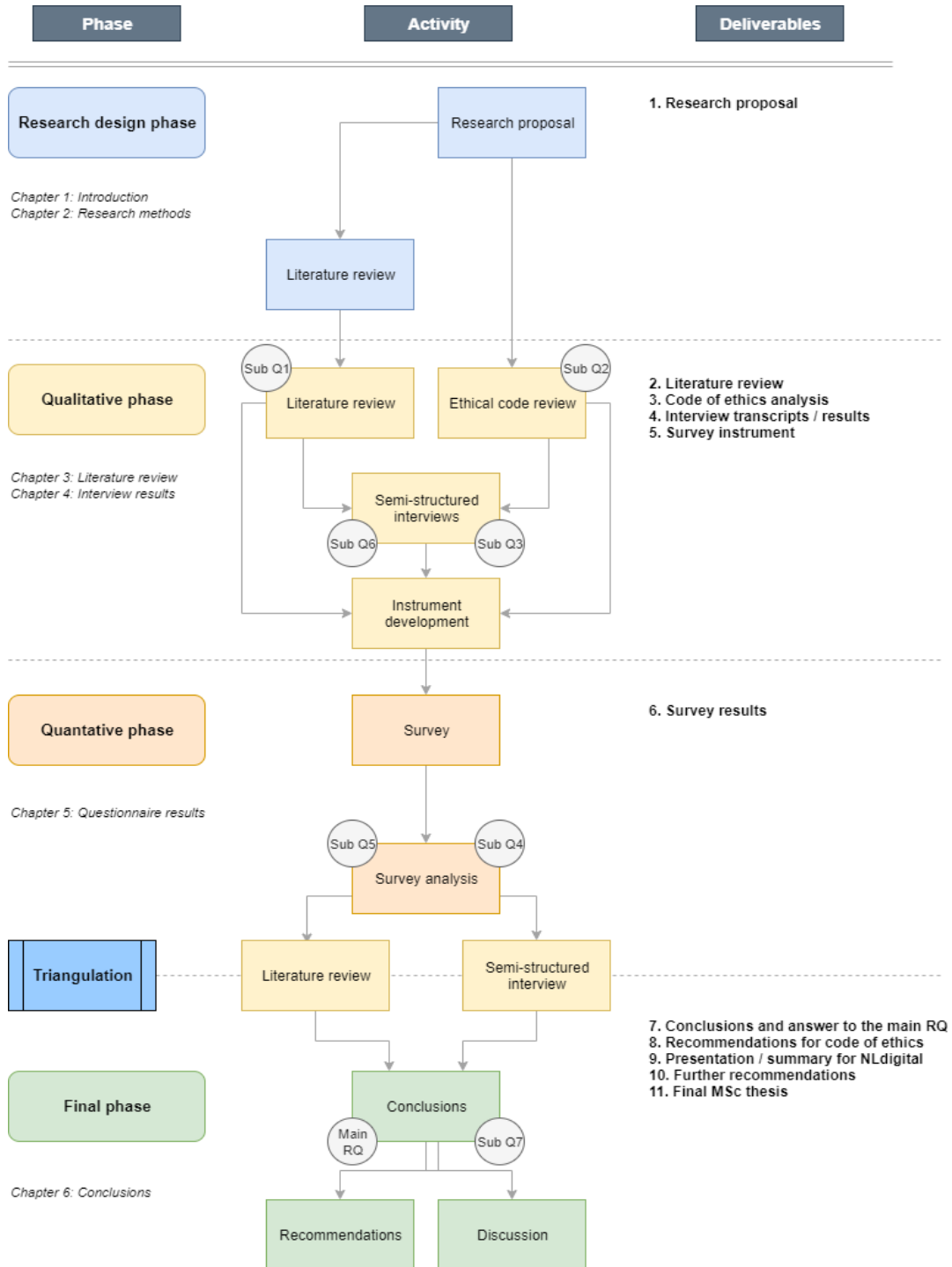


Figure 1: Research flow diagram

2.2 Research design

This section describes the design and structure of each of the data collection methods. Section 2.2.1 delineates the procedures for the interviews and 2.2.2 lays out the procedures for the questionnaire. The book of Sekaran and Bougie (2016) is used as a guideline for both methods.

2.2.1 Interviews

As discussed in the previous sections, semi-structured interviews are used as part of the qualitative data gathering, and to develop the questionnaire instrument. This section discusses the scope and protocol for the interviews and describes the criteria for the interviewee selection. The results of these interviews are discussed in chapter 4.

2.2.1.1 *Scope definition*

The scope for the interviews is focused on, but not limited to member companies of NLdigital. There is the assumption that these companies represent a cross-section of the part of the ICT industry. Furthermore, a precondition for all the respondents is to make use of artificial intelligence either as a product or as a service that they offer. These preconditions are chosen because: 1) the scope of the interviews is limited to businesses in the ICT industry working with AI, 2) these companies should experience ethical challenges with AI, 3) these companies should be familiar with NLdigital's code of ethics. In the end, five-member companies and one non-member company in the ICT-industry that are working with AI were interviewed. This research recognizes the limitation of not including more ICT companies and experts from outside of the member base in the sample. Despite this limitation, I believe that these companies are representative of the general sentiment regarding the use and implementation of AI in the Dutch ICT industry.

2.2.1.2 *Interview protocol*

An interview protocol consisting of a number of questions is formulated. These questions should be answered throughout the interview. However, the order of the questions in the protocol does not necessarily need to be followed throughout the interview. This means that if it is logical in the conversation to ask a question earlier than it is mentioned in the protocol this forms no problem for the quality of the qualitative data. Furthermore, the protocol allows for further exploration of the topics that interviewees raise. The interview protocol (see Appendix A) consists of an introduction, permission to record and a list of 8 main questions to be asked and discussed. The questions aim to understand the ethical challenges that these companies encounter and the way in which they deal with them. The questions also aim to discover any opportunities for improvement in measures for dealing with ethics and the current code of ethics. The protocol was formulated based on the sub-research questions, the literature review, and feedback from different experts during the early stages of the project. Important concepts from the academic literature that are used include the ethical challenges with AI that are recognized as most important (e.g., fairness, transparency, and privacy), and a framework for good ethical governance of AI (Winfield & Jirotko, 2018). The interview protocol was finetuned by requesting feedback from experts prior to the interviews.

All interviews are conducted in a face to face setting to capture any nonverbal clues and have a more engaging conversation. This is important because of the semi-structured nature of the interviews. The entire interviews are recorded and transcribed into text to cite the interviewees accurately and allow for further analysis in atlas.ti.

2.2.1.3 Interviewee selection

For the interviewee selection, this study uses a combination of selective and convenience sampling. The sample is taken from active and new members of NLdigital that are known to be working on artificial intelligence. The advantage of this sampling method is that it's easy and time effective because these participants are easy to reach and known to be willing to co-operate. The advantage of selecting specific companies and people within those companies out of the convenience sample is that it ensures detailed and in-depth responses on the topics of interest. The disadvantage and limitation of potential biases due to this choice of sampling is recognized but mitigated by not interviewing the 'contact-person' of NLdigital, but rather the people from those organizations that are not in direct or frequent contact with NLdigital.

For the selection of organizations, 2 large enterprises (i.e., more than 250 employees) and 4 small and medium enterprises (SMEs) (i.e., less than 250 employees) were chosen. More SMEs were included because of the hypothesis that these were underrepresented in the design of the code of ethics by NLdigital, and because of the hypothesis that these organizations have less ethical measures in place than larger organizations. Two large organizations are still included to show this distinction and map out the mechanisms that such companies use. For the selection of the interviewees within those organizations, employees that are actively working with AI are preferred. The requirements for interviewees are:

- Uses some form of AI or machine learning in his or her work, either applying existing models and systems to their own products or services or develop his or her own systems.
- Is aware of practical and ethical challenges with AI and has experience in hands-on handling or explaining those.

Table 2 shows the complete list of candidates that were eventually interviewed for this research.

Table 2: Interview respondents

Interviewee	Company	Footprint	Size	Member	Duration	Date
IV1	n.a.	Global	Large	Yes	45 min	30-10-2019
IV2	n.a.	International (EU)	Small	Yes	40 min	31-10-2019
IV3	n.a.	National (NL)	Small	Yes	40 min	1-11-2019
IV4	n.a.	Global	Large	Yes	59 min	1-11-2019
IV5	n.a.	Global	Medium	Yes	57 min	5-11-2019
IV6	n.a.	National (NL)	Small	No	34 min	6-11-2019

2.2.1.4 Reliability of the coding process

The semi-structured interview coding presents the typical limitations of qualitative text analyses such as that of subjective bias in the coding process. Following best practices for coding in-depth semi-structured interviews, we mitigated this limitation by validating our coding strategy by an independent external coder (Campbell, Quincy, Osserman, & Pedersen, 2013). The level of similarity in coding is used to assess the interrater reliability of the coding. To determine the reliability of the coding process, coding 10% of the interview transcripts is generally accepted as sufficient (Campbell et al., 2013). For the purpose of this research, the external coder was asked

to code the first two interview transcripts in Atlas.ti. This amounts to 33% of the total amount of interview transcripts or 30% (14 pages) of the transcribed pages. This is well over the accepted norm of at least 10% of the transcripts and 5 transcribed pages. The coding scheme in our study was generated mostly in a deductive way, meaning that the information and knowledge acquired during the ethics and AI literature review forms the basis of the generated codes. This guided what information and concepts we were looking for in the interviews. As a result of this deductive approach to coding, it is assumed that an external coder cannot transcribe the interviews in a similar way without in-depth prior knowledge on the subject. Therefore, a brief training module consisting of an outline of the core concepts, the concepts of interest for this study, and a few examples of the codes were shown to the external coder.

The most common statistic to calculate intercoder reliability in content analysis is Krippendorff's alpha (Krippendorff, 2004). However, the use of Krippendorff's α assumes that both coders have the same qualifications and level of knowledge of the subject of interest. This is a criterion not met in this study, as is frequently the case in researches that use in-depth semi-structured interviews (Campbell et al., 2013). In semi-structured interviews the chance that two coders code the same quote of an interviewee with the same code is therefore considered negligible, resulting in a low α (Mouter & Vonk Noordegraaf, 2012). For our study, a more useful metric to demonstrate the intercoder reliability is Holsti's coefficient. Holsti's coefficient analyzes the degree of agreement between the coders of that a certain quote is marked as a problem. Holsti's coefficient of the interview coding for this research was determined to be 0.85 (see equation below).

$$\text{Holsti's coefficient} = \frac{2 * \sum_{n=0}^2 (C1(n), C2(n))}{N1(n) + N2(n)}$$

where $N1(1) = 42$, $N1(2) = 29$, $N2(1) = 41$, $N2(2) = 37$

and $C1(1), C2(1) = 70$, $C2(1), C2(2) = 56$

Holsti's coefficients of at least .80 are widely accepted as reliable results for most academic research (Mouter & Vonk Noordegraaf, 2012). In consideration of the abovementioned argumentation and a Holsti's coefficient of 0.85, we consider the coding and analysis of the semi-structured interviews in this study to be reliable.

2.2.2 Questionnaire

This section describes the design process of the questionnaire that is used to gather the descriptive data in this research. Sub-section 2.2.2.1 describes the survey methodology, sub-section 2.2.2.2 describes the question design, sub-section 2.2.2.3 lays out the sampling approach, and finally, sub-section 2.2.2.4 delineates the reliability of the sampling approach and questionnaire.

2.2.2.1 Methodology

The type of survey that is chosen for this research is a questionnaire. This is a valuable data collection method for collecting descriptive and exploratory information, as is the aim of this research (Sekaran & Bougie, 2016, p. 147). The questionnaire will be electronically distributed and e-mailed to all the members of NLdigital. An electronic questionnaire is a time-effective way to describe an as-is situation for a larger sample than with interviews. Moreover, it is easy to administer, and respondents can answer questions at their own convenient times and pace.

However, the method also knows some disadvantages and limitations (Sekaran & Bougie, 2016, p. 148). First, a high non-response rate is much more likely than with interviews. This risk is mitigated by scheduling a reminder of the survey after a week, creating awareness and traffic around the survey on NLdigital's LinkedIn channel, and having a back-up plan of a different ICT-company network that can be tapped into. Another way in which a high response rate is stimulated is by incentivizing respondents with the provision of a professional sheet with the key results of the survey and interview after the results have been analyzed. Another limitation of an electronic questionnaire is that questions cannot be clarified, this is a problem if some respondents do not understand a certain question or word. This is mitigated by receiving feedback on the survey from experts and testing the questionnaire for any unclear questions and errors before sending it out. Finally, another limitation of using an electronic questionnaire is that respondents may give biased responses. There is a risk that respondents give socially acceptable answers and pretend to be more ethical in their AI practices than they actually are. To limit the risk of biased responses the respondent will be told at the start of the survey that his/her responses will be anonymized.

The questionnaire is designed to map out the AI-landscape, determine the importance of the different ethical challenges with AI as perceived by the respondents, map out the mechanisms and measures to increase or enforce ethical behavior that companies have in place, and determine the adoption by the industry of NLdigital's code of ethics. Because the questionnaire aims to map out these aspects, the nature of the questionnaire is descriptive, and its primary goal is to collect descriptive data with regards to the abovementioned concepts. This fits within the framework of the research questions that guide this research (section 1.4) and will contribute to the answering of the main research question: 'Which ethical challenges are currently missing in the current code of ethics from NLdigital and are fundamental for guiding responsible research and development in the use and application of AI-systems for the Dutch ICT-industry?'

2.2.2.2 Question design

As previously indicated in section 2.1, the survey instrument is developed based on the results of the interviews (chapter 4) and the ethical building blocks that were identified and delineated throughout the literature review (chapter 3). The literature review is used in a similar way as in the creation of the interview protocol, namely, to motivate the inclusion of key ethical building blocks such as the five key ethical challenges with AI and the framework for solid ethical governance. The interviews form the key input for the survey development. The majority of the questions are identical in objective as those in the survey (i.e., pressing ethical challenges, mechanisms, pillars of ethical governance, and feedback on the code of ethics) because the latter proved to be relevant and informing during the interview process. For the survey, the concepts and wording that make up the questions are polished using the experiences and feedback gained throughout the interviews. This results in more precise and targeted survey questions. Some new dimensions were discovered during the interviews such as the distinction between AI for internal processes, products, and services, the way to ask about feedback on the code of ethics, and the difficulty of ranking the impact of the five key ethical challenges on the spot. These dimensions and findings are all considered in the design of the survey instrument. The final questionnaire is displayed in Appendix C.

The scope of this research is broad as it intends to describe a wide range of AI and ethical practices in the Dutch ICT-industry. This also means that a wide range of data will be collected.

The first four questions of the questionnaire aimed to collect demographic information about the different organizations. This information is used to check how footprint, company size, and location of headquarters influences the occurrence of ethical challenges and the presence of ethical mechanisms and pillars. The company name is asked to filter duplicate responses from the same organizations, but these names are omitted in the publication of this work. The second section of the questionnaire aims to map out the current AI-landscape by enquiring into the stage, use-case, development process, and target market of companies' AI-enabled solutions. These questions are all multiple-choice questions with the possibility to add an answer that is not listed. The next section deals with the ethical challenges that organizations encounter with AI. Respondents are asked to rate the importance of different challenges for them and list their most pressing challenges. This is done using a 5-point asymmetric Likert scale. A 5-point scale, as opposed to a 3-point scale, is chosen because of the additional value it provides in understanding the value that a respondent assigns to an answer. The fourth section of the questionnaire has the objective to describe the mechanisms to enforce ethical behavior and the pillars of ethical governance that organizations have in place. The multiple-choice answers for the mechanisms are constructed with the answers provided in the interviews. The final section of the survey asks people about their familiarity with and use of the code of ethics to determine the code's current adoption and impact. Respondents are also asked to rate the extent to which their current practices comply with the principles in NLdigital's code of ethics. Finally, respondents are given the opportunity to provide feedback on the questionnaire.

The abovementioned rationale behind the design and structure of the questions is focused on answering exploratory research questions rather than quantitative ones in which one tests relationships between variables. This is a common approach in constructing surveys for descriptive research. The questionnaire will for a large part answer sub-research question 3: 'Which ethical challenges are most pressing to Dutch ICT-businesses in the development and application of AI-systems?'. Furthermore, the results of the questionnaire will together with the results of the interview and theoretical support from the literature answer sub-research question 5, 6, and 7: 'What concrete measures do Dutch ICT-businesses currently use to deal with the ethical aspects of developing and applying AI-systems?', 'What are the common concerns that are uttered by the Dutch ICT-community about the current code of ethics offered by NLdigital?', 'What concrete measures are missing that would assist small and medium-sized companies in dealing with ethical challenges, and how does a code of ethics play a role here?'.

2.2.2.3 Sampling approach

The targeted population of this study is companies and organizations active in the Dutch ICT-sector. The member directory of NLdigital is chosen as the main sample frame because the subjects in this sample match the population exactly. This means that the members of NLdigital that will make up the sample are all part of the targeted population. This does signify that a type of non-probability sampling is used. The members of NLdigital form a convenience sample of the population. They are sampled because they are conveniently available to provide responses to the questionnaire. This type of sampling is common in exploratory research because of its ability to gather a lot of data within a limited time frame and with limited resources. The limitation on the generalizability to the entire population is recognized but is believed to form a minor challenge in this research as NLdigital's members are known to be a representative cross-section of the ICT-

industry. In total, the questionnaire was sent out to over 600 member companies of NLdigital and around 150 members of the KNVI. The focus was to get the highest response of companies within the sector that are working with AI but also gives companies not (yet) working with AI-systems the opportunity to respond to the questionnaire.

2.2.2.4 Sample merging and reliability

Despite numerous efforts to increase the response rate of the questionnaire among NLdigital's members, a response rate of around 5% was too low to establish significance in the results that result in a commonly acceptable error margin (i.e., 5-10%). It was therefore decided to extend the sample beyond the original sample frame of NLdigital's members. An additional list of respondents comes from the KNVI (Koninklijke Nederlandse Vereniging van Informatie professionals). The KNVI is an employee organization and functions as a platform for professionals in the ICT-sector in the Netherlands. This means they fit the same population as the initial sample frame, i.e., companies and organizations active in the Dutch ICT-sector. Although the members of the KNVI fit the same population, the implications for the reliability of the results of questions 14 and 15 (appendix C) should be recognized because we cannot assume members of the KNVI to be familiar with NLdigital's code of ethics.

In the end, the survey was sent out to a total of over 750 people. Despite numerous efforts to surpass the typical response rate of NLdigital's surveys (i.e. between 3% and 5%), our efforts resulted in a response rate of around 6%. The electronic questionnaire recorded a total of 46 respondents from NLdigital (37) and the KNVI (9). However, after removing incomplete responses, 34 valid responses were recorded on 40 variables (appendix D). For the calculation of the margin of error of our results, we use a total population of 50.000 ICT-companies (ABN-AMRO, 2019). Table 3 displays the margin of error of our results on different confidence intervals. These margins of error fall outside the generally accepted range of 5-10% and pose limitations on the confidence of the generalizability of the results. This effect is partially mitigated, however, by the indication of similar results from both the interviews and the questionnaire, and results that are in line with what we can reasonably expect from literature and discussions with different experts.

Table 3: Margin of error on different confidence intervals of the electronic questionnaire

Confidence level	Margin of error
95%	17%
90%	14%
85%	12%
80%	11%

2.3 Conclusion

Due to the exploratory and descriptive nature of this research, a mixed methodology approach is identified as the most suitable research approach for answering the main research question and the sub research questions. More specifically, this research will follow the exploratory sequential mixed methodology, which is typically recognized by a qualitative research phase followed by a

quantitative research phase. In this research, an extensive literature review and a set of six semi-structured interviews will answer sub-questions 1, 2, and 3:

1. What are the chief ethical aspects in the research, development, and application of AI-systems for businesses in the Netherlands that are identified in the specialized literature?
2. What ethical aspects in NLdigital's code of ethics are missing that can be identified by comparing it to other codes of ethics?
3. How do companies in the Dutch ICT-industry perceive the ethical challenges of AI?

The literature review and interviews are followed by a questionnaire that is sent out to all the members of NLdigital. The questionnaire allows this research to arrive at descriptive conclusions for a large part of the ICT-sector. It will also partially answer sub-research question 4, 5 and 6:

4. Which ethical challenges are most pressing to Dutch ICT-businesses in the development and application of AI-systems?
5. What concrete measures do Dutch ICT-businesses currently use to deal with the ethical aspects of developing and applying AI-systems?
6. What are the common concerns that are uttered by the Dutch ICT-community about the current code of ethics offered by NLdigital?

The last sub-research question (i.e., What concrete measures are missing that would assist small and medium-sized companies in dealing with ethical challenges, and how does a code of ethics play a role here?), as well as the main research question, will be answered by triangulating the three methodologies. The analysis of NLdigital's code of ethics will identify areas for improvement among the eight defined principles. The literature review will motivate why it is important to include, remove, or adjust certain principles. Finally, the interviews and questionnaire will demonstrate that these are important challenges that companies in the industry have to deal with and need guidance with. Bringing these three angles together will answer the main research question of this research: *'Which ethical challenges are currently missing in the current code of ethics from NLdigital and are fundamental for guiding responsible research and development in the use and application of AI-systems for the Dutch ICT-industry?'*. Answering this question is important because it adds a new practical applied perspective to the body of research on AI ethics. This study contributes with a descriptive study of the ICT-sector, which is the leading industry with regards to the development and implementation of AI-systems. On top of research on ethical challenges in the industry, this work provides descriptive insight into the way in which AI is currently governed in organizations and what tools businesses use to encourage ethical behavior with artificial intelligence. This is a topic that is currently mainly discussed on an abstract level rather than from an applied perspective. This work fills this knowledge gap by mapping out the concrete measures that are used in the industry and how small and medium-sized enterprises can benefit from a code of ethics to govern AI-ethics.

3. Literature review of artificial intelligence and ethics

This chapter functions as the first part of the qualitative research phase. It will build upon the identified research problem to build a contextual framework in the technical domain (section 3.1) and identify the key ethical challenges with AI in the theoretical domain (section 3.2). This section also lays the basis for codes of ethics as a concept and discusses different ways organizations anticipate and solve them to provide direction for the interview. Afterward, section 3.3 analyzes different codes of ethics and guidelines to find common concepts among these. The chapter will conclude (section 3.4) with the key ethical concepts that together with the interview results (section 4) will form the key ethical building blocks and the foundation for the questionnaire. The delineation of the ethical challenges will together with the interviews answer sub-research questions 1 and 2. The first sub research question aims to find out what the ethical challenges are that companies in the Netherlands must deal with in the development and application of AI, and the second one how companies in the industry perceive ethical challenges of AI. This fits the objective of this study to map out the key ethical challenges with AI in the industry

3.1 Technical domain

This section will unfold the most relevant technical aspects of artificial intelligence. Artificial intelligence as a theoretical concept in general (section 3.1.1) is described, as well as its subsets machine learning and deep learning (section 3.1.2). Finally, the difference between the design and application of AI systems is delineated as well as the key technical challenges for the industry (section 3.1.3).

3.1.1 Artificial intelligence

Present-day, artificial intelligence is one of the most popular topics in businesses, government, institutions (e.g. hospitals) and academia. The scientific field of AI has seen more than a nine-fold growth in the annual publishing rate of academic papers in the last two decades (Shoham et al., 2018). This section reviews the academic history (section 3.1.1.1) of AI briefly, provides a working definition (section 3.1.1.2), and discusses AI as the umbrella term of other self-learning techniques (section 3.1.1.3). In section 3.1.1.4 promising applications of AI will be explored and in section 3.1.3.3 the most pressing challenges to its further development and acceptance into society are laid out.

3.1.1.1 History

The discipline of artificial intelligence has been around for much longer than the last two 'popular' decades of AI. The first AI system was invented at Carnegie Mellon University by Newell, Simon, and Shaw in 1955 (Flasiński, 2016, p. 4). The system, named 'Logic Theorist', was able to prove a series of fundamental mathematical theorems. A year later, John McCarthy, one of the other founding fathers of AI, organized the Dartmouth conference which is widely recognized as the start of AI as an academic discipline. In the years that followed, until the late 90s, interest, and funding in AI was unstable to the extent that the field went through two 'AI winters' in which very

limited research funding was available. In 1997, AI research gained serious momentum again after a Russian chess champion was defeated by IBM's Deep Blue system (Lewis, 2014). The next sensational breakthrough for artificial intelligence came in 2016 when Alpha Go defeated a Go champion in the game Go, which is widely seen as one of the most difficult strategy games in the world (S. Singh, Okun, & Jackson, 2017). Unlike Deep Blue that used search trees to master chess, Alpha Go used machine learning and reinforcement learning to teach itself how to play. This showed for the first time that computers can learn through practice and experience like humans do, and thereby marked the start of self-learning artificial intelligence that we see used in businesses today.

3.1.1.2 Definition

Today, many different definitions of Artificial Intelligence exist throughout the academic and the business community. Legg and Hutter (2007) present a wide collection of definitions of intelligence from multiple perspectives. These definitions are a collection of over 80 definitions from encyclopedias, dictionaries, psychologists and AI researchers. Nevertheless, they observe three common features in the definitions that they reviewed in their research. A new definition that incorporates these three definitions is proposed: "intelligence measures an agent's ability to achieve goals in a wide range of environments" (Legg & Hutter, 2007, p. 9). The first definition of AI in an academic field stems from McCarthy who defined it as "*the science and engineering of making intelligent machines*" (Rajaraman, 2014, p. 206). A more modern definition used in contemporary textbooks takes this definition a step further by adding that an intelligent machine always wants to maximize an objective (Russell & Norvig, 2016). Russell and Norvig (2016) include the notion that intelligent agents perceive their environment and take actions that maximize its chances of success.

Different concepts of artificial intelligence have been mixed in contemporary culture and debates. Researchers, therefore, choose to draw a distinction between strong AI or Artificial General Intelligence (AGI) and weak or narrow AI (Yao et al., 2018, pp. 8-9). AGI refers to systems that have the capability to learn and extract knowledge from one domain and apply it to another domain. These systems are generally associated with trying to replicate human cognitive capabilities to understand the world as well or better than we do. Although organizations are working on such systems, AGI does not exist yet. Any AI system in use now would, therefore, be labeled as weak or narrow AI. This term refers to systems that are specifically designed for one task (Yao et al., 2018, pp. 8-9). Kaplan and Haenlein (2019) take the definition of narrow AI further by adding how an AI-system carries out a task and achieves its objectives. They propose: "AI is a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation (Kaplan & Haenlein, 2019, p. 17).

More practical definitions emerge from the industry rather than academic literature. Giant in the industry, Intel, defines AI as "*an umbrella term for a program that can sense, reason, act and adapt*" (N. Singh, 2016, p. 1). NLdigital deliberately used a broad definition of AI in the development of the ethical code to make it as broadly applicable. They defined AI as everything that touches self-learning or self-reasoning systems (van der Beek, 2019). This definition, however, is not narrow enough to adopt as a working definition for this research as it could also include biological self-learning and reasoning systems such as humans.

For the purpose of this research, some elements of the presented definitions are used, and some other constraints are added to make it fit the scope of this research. For this study, AI will be considered as machines that are programmed to interpret external data, learn from such data, and use the learnings to successfully carry out a single task. These 'narrow' systems carry out this task within a pre-established range, even if they appear to be operating in a much more sophisticated way than that.

3.1.1.3 Artificial intelligence as an umbrella term

Artificial intelligence as a discipline has been around since the 1950s (Flasiński, 2016). Over the course of the last few decades, new AI techniques have been developed. Machine learning is by far the most used AI technique used in contemporary AI applications and is widely recognized as a subset of the umbrella term artificial intelligence as is displayed in Figure 2 (Arel, Rose, & Karnowski, 2010; N. Singh, 2016). Machine learning is discussed more elaborately in section 3.1.2. Another subdivision can be made for deep learning, a more recent technique that is seen as a sub-discipline of machine learning (N. Singh, 2016; Yao et al., 2018, pp. 9-14).

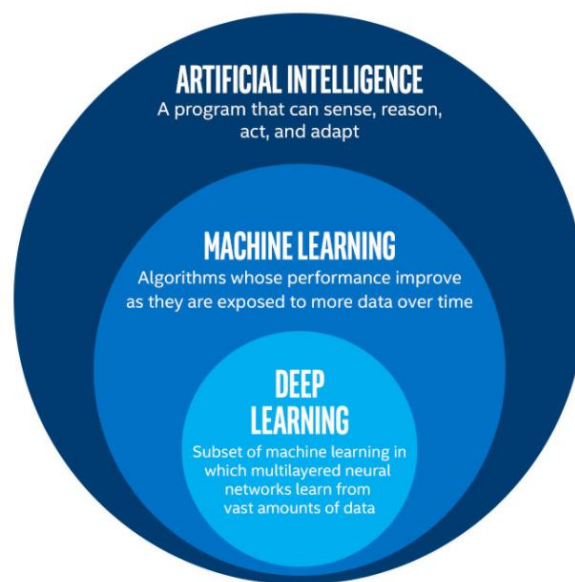


Figure 2: Artificial intelligence subdivision, adapted from: (N. Singh, 2016)

3.1.1.4 Possibilities

Disruptive technologies such as autonomous vehicles and robots are some of the most debated AI-enabled technology topics discussed in society today. Self-driving cars can vastly reduce the number of accidents and casualties in traffic, save time, and open the possibility for completely new mobility concepts. Robots could take over repetitive and dangerous tasks from humans. Yao et al. (2018, p. 31) describe some fields through which AI can bring significant improvements to the well-being of people in societies all around the world. They observe that artificial intelligence systems could bring rapid innovation and progress to microfinance, social justice, and medical diagnosis. Makridakis (2017) examines the impact artificial intelligence will have by contrasting it with the industrial and digital revolutions as analogous inventions. He concludes that a so-called 'AI revolution' will have a larger impact on society and industry than the industrial and digital revolution combined. Makridakis also suggests that significant competitive advantage can be

gained by companies that are willing to take risks and be pioneers in the innovation of AI technologies (Makridakis, 2017). Currently, the ICT industry forms the backbone for much of the development of AI applications. Both start-ups and large incumbents in the software space are leading the AI revolution and growing their AI practices rapidly.

For the Netherlands specifically, more artificial intelligence adoption will boost its economy. Research by Accenture shows that the Netherlands can grow its labor productivity by 27% more compared to the baseline that is expected by 2035 if it were to absorb more AI into the economy. As a result, the Dutch economy could realize an annual growth rate of 3.2% in this scenario, compared to the otherwise projected 1.6% (Purdy & Daugherty, 2016).

3.1.2 Machine learning

Artificial intelligence and machine learning are often used interchangeably. Even though almost all of modern AI applications are enabled by machine learning algorithms, they are not the same. However, most of the interviewees and survey respondents in this research will use a form of machine learning. Therefore, it is important to establish the context of machine learning as well. This section describes machine learning as a subset of artificial intelligence. In section 3.1.2.1 different definitions from academics and industry are contrasted and in section 3.1.2.2 the four different learning methods of machine learning are laid out. Then, in section 3.1.2.3 the machine learning life cycle is broken down and finally in section 3.1.2.4 deep learning is briefly discussed.

3.1.2.1 Definition

Faggella (2019) created a definition of machine learning (ML) by aggregating definitions and reviews from multiple experts in the field. In his online article, he proposes the following definition: “*Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in an autonomous fashion, by feeding them data and information in the form of observations and real-world interactions*”. Yao et al. (2018, pp. 12-13) add to the notion that computers can learn, that in machine learning computers are not explicitly programmed. Raschka (2015, p. 1) uses a wider definition and considers machine learning to be makings sense of data through algorithms.

SAS, a large software company and statistical analytics application, provides a clear distinction between AI and ML. Where AI is concerned with copying human abilities to machines, ML is an approach to implement AI by training machines on how to learn (SAS, 2019).

3.1.2.2 Learning methods

Computers can be trained to learn through different methods. Machine learning methods can be subdivided into a series of four different learning methods (Yao et al., 2018, pp. 9-14). Table 4 provides an overview of the different learning methods and their corresponding algorithms as defined by Yao et al. (2018, pp. 9-14) along with some of the most prominent fields of application.

Table 4: Machine learning methods

Method	Learning	Algorithms	Application examples
Supervised learning	Learns rules by training it with labeled data in which inputs and outputs are paired.	Regression Classification	Image recognition, Forecasting
Unsupervised learning	Searching categories and structure in unstructured and unlabeled data	Clustering Association	Customer segmentation
Semi-supervised learning	Learns by training with some labeled data and a large amount of unlabeled data.	Classification Clustering	Recommender system
Reinforcement learning	Learn by trial and error to reach a maximum reward.		Robotics, Games

3.1.2.3 Machine learning life cycle

For the practical application of machine learning models, businesses often speak about the machine learning life cycle. This cyclical model can be used to guide machine learning and data science projects in business and consists of four main steps (Etaati, 2019).

- I. Business understanding
- II. Data acquisition and understanding
 - a. Data collection
 - b. Feature selection
 - c. Data wrangling
- III. Modeling
 - a. Model selection
 - b. Split data set
 - c. Train model
- IV. Deployment
 - a. Evaluating the model
 - b. Monitoring model

In her book, Etaati (2019) elaborates on the phases of the machine learning life cycle.

3.1.2.4 Deep learning

The last learning method and a subset of machine learning is deep learning. Because of the high performance and accuracy of deep learning, this is the preferred method by technology giants for large scale applications, e.g. Google's translation engine and Microsoft's speech recognition engine (Yao et al., 2018, pp. 15-16).

An extensive definition comes from Deng and Yu (2014) who analyzed common elements among a set of five definitions on deep learning. Two important elements were found (Deng & Yu, 2014, p. 201):

- I. “models consisting of multiple layers or stages of nonlinear information processing”
- II. “methods for supervised or unsupervised learning of feature representation at successively higher, more abstract layers.”

On top of the notion of multiple (abstract) layers, Yao et al. (2018, pp. 15-16) describe that deep learning algorithms use artificial neural networks to perform classification tasks. LeCun, Bengio, and Hinton (2015) describe in their article ‘Deep Learning’ in magazine Nature the fundamental difference with other machine learning techniques. Where other machine learning techniques require a designer of a system to perform many different actions and feature extractions to classify data, deep learning allows a system to automatically determine useful representations to classify raw data. Although deep learning can be vastly more accurate than conventional machine learning techniques, it also requires very large amounts of data and is computationally intensive (LeCun et al., 2015).

3.1.3 Design and implementation of Artificial intelligence

This section describes the different ways in which businesses can work with artificial intelligence. Section 3.1.3.1 lays out what it means for businesses to design artificial intelligence and section 3.1.3.2 describes what it means to implement artificial intelligence into operations, products, and other artifacts. Finally, section 3.1.3.3 delineates the technical challenges with AI for businesses.

3.1.3.1 *Design*

This research considers the design of AI to be the creation of sophisticated AI algorithms or software programs. Typically, developers will do this by coding in a programming language of their or their organization's choice, e.g. Python, Java, C++ or LISP. Most often, companies will use existing algorithms, frameworks, tools and libraries to tailor solutions to their own needs. Some of these tools are open source libraries licensed by large incumbents such as Google's Tensorflow, some by academic institutions such as MIT's Keras, and others have free software licenses (BSD) such as Torch and Scikit-learn.

One of the most recent advancements in widespread machine learning for businesses is that of ‘Machine Learning as a Service’ or MLaaS. Designing your own AI or machine learning system is a very specialized task and requires a vast amount of ML knowledge and experience. This creates barriers for companies with no access to AI talent or insufficient resources to use homegrown artificial intelligence in their practices. Other companies have seized this opportunity by creating the machine learning as a service business model (Hunt, Song, Shokri, Shmatikov, & Witchel, 2018). Google, Microsoft, IBM, Amazon, and a few smaller competitors provide sophisticated machine learning platforms for businesses that do not have the expertise or will to code their own machine learning models. Moreover, MLaaS is a relatively affordable way for businesses to reap the benefits of machine learning. Although MLaaS is an attractive offering for many businesses, Hunt et al. (2018) observe two important barriers for adoption, i.e. privacy and confidentiality. Users of MLaaS will have to provide the data that the ML model will be trained on to the service provider which might cause them to share confidential information or lose their competitive edge. Furthermore, this also raises some privacy concerns in terms of unknowingly or perhaps unwillingly sharing data of their users. In the end, the true winners of the recent trend in MLaaS are the large service providers who are gaining very large amounts of data on which they can further train and develop their models and algorithms.

3.1.3.2 *Application*

Many of the companies in the Dutch ICT industry will be applying and modifying existing artificial intelligence algorithms or systems to their specific product or solution. They will use services such as IBM Watson, Google Cloud AI, Microsoft Azure machine learning or Amazon Machine Learning Services to enhance their businesses processes or create novel products. The assumption should, therefore, be made that companies using such services will not always fully understand the system they are using and cannot influence the design of such complex AI. They are handling what is in their own sphere of influence, which is the process of applying the technology for their own purposes. Applications of artificial intelligence in the ICT industry are very wide-ranging. This has to do with the nature of the industry, which is providing information technologies to businesses in other industries. Information technologies are almost all based on data, they are often systems that create, store, send or use data. Applications of AI in information technology can range from AI-assisted and data-based decision making for various companies to the creation of automated systems such as chatbots. For businesses active in telecommunications, artificial intelligence is a very important enabler of new technologies such as 5G technology. In 5G networks alone, AI is used in subfields such as “radio resource management, mobility management, general management and orchestration, and service provisioning management” (Li et al., 2017, p. 2).

3.1.3.3 *Challenges*

Even though artificial intelligence offers many opportunities for the advancement of businesses and society, further progress faces some significant challenges. First, there are technical challenges that inhibit the full potential of AI. Immense computational power is needed to run the state-of-the-art machine learning and deep learning algorithms, especially because these systems are expected to calculate almost in real-time (Adixon, 2019). Current systems with massive processing power such as cloud computing and parallel processing are reaching their current limits. Innovation in this space is needed for the further development of AI. Another technical challenge for AI is that most AI systems are currently built for one specific task, this is called narrow or weak AI (Yao et al., 2018). Generalized AI, systems that can apply their ‘learnings’ to other tasks like humans, have yet to be developed. Lastly, well-labeled data sets that are large enough for creating accurate AI applications are scarce in most businesses and industries.

These challenges will also form barriers for the effectiveness and further development of AI in some businesses in the Dutch ICT-sector. First, the quality of the output of an AI system is dependent on the quality of the data that you train the algorithm on. This provides a competitive advantage for the large incumbent ICT companies, that generally have better access to very large quantities of data and employ people and infrastructures specifically for maintaining the quality of their data sets. This gives them an increased advantage over their smaller competitors who lack this kind of data assets (Ransbotham, Kiron, Gerbert, & Reeves, 2017). Second, the complexity of sophisticated AI techniques and the scarcity of talent in AI forms a barrier for further development. Ransbotham et al. (2017) found that for pioneers the biggest challenge is to educate and acquire AI talent. Training and building algorithms and integrating them with business applications is a highly specialized skill. For businesses that are more passive in the deployment of AI, this is also an important barrier, moreover, these companies are often not even aware of the need for developing AI talent.

Another set of challenges arises in the societal field. People oftentimes view AI as a black box, a process in which some ‘unexplainable’ decision is made. Many people do not trust AI systems because they do not understand how the system reached a certain outcome (Manyika & Bughin, 2018). Second, according to McKinsey research, employment for certain jobs will decline, while for others their current work may change significantly as a result of the effects of AI in automation (Manyika & Bughin, 2018). It will be an important challenge for developers, policymakers and the people, in general, to deal with the changing job landscape due to automation. The biggest challenges of all, perhaps, are the various ethical challenges that artificial intelligence poses. Data privacy, inequality, accountability, biases and many other challenges will need to be resolved for the sustainable development and acceptance of AI in the future (Baylé, 2019).

3.1.3.4 Implications for the rest of this research

This sub-chapter (section 3.1) has laid out the technical context of this research. Specifically, we have looked at artificial intelligence being the umbrella term for machines that are programmed to interpret external data, learn from such data, and use the learnings to successfully carry out a task. We found that these systems are currently still very ‘narrow’, and mostly carry out a single task within a pre-established range, even if they appear to be operating in a much more sophisticated way than that. Then, we discussed machine learning in more detail. Machine learning is the umbrella term for what most contemporary AI-systems are powered by. Based on this literature, and some exploratory conversations with businesses in the ICT-industry, we expect most, if not all, businesses in the ICT-sector to be working with machine learning if they are using AI-systems. Businesses can use machine-learning algorithms to optimize internal processes, enhance their products, or deliver services. In sub-section 3.1.3.2, for example, we described how telecommunication providers can benefit from machine learning algorithms for 5G communication technology. For the information technology division of the ICT-industry, machine learning is expected to be used as a service to successfully build systems around (unused) data of their customers.

The literature review in this sub-chapter was done to provide the technical context and understanding for the ethics part of the literature review, interviews and questionnaire. Having a thorough understanding of the differences between AI and ML, the possibilities of AI, and the limitations, will allow us to assess the ethical implications of it in an objective and informed way. The next sub-chapter will explore these ethical implications in depth by delineating the ethical aspects of AI for business by describing the different ethical challenges and reviewing the relevant concepts and theories.

3.2 Theoretical domain

While the first part of this literature review delineated the technical domain of this research, this chapter aims to create a theoretical understanding of ethics and AI ethics. First, section 3.2.1 lays out the fundamental ethical theories. This is important for understanding how businesses in the industry perceive ethical challenges and what justifies their way of handling these challenges. Second, prominent literature on the ethics of artificial intelligence are discussed and important ethical challenges with AI are identified in section 3.2.2. This is essential to recognize the challenges that businesses in the ICT-sector are dealing with. This section will be important to answer the first two sub research questions (i.e., What are the chief ethical aspects in the

research, development, and application of AI-systems for businesses in the Netherlands that are identified in the specialized literature? What ethical aspects in NLdigital's code of ethics are missing that can be identified by comparing it to other codes of ethics?) and inform the content of the interviews and questionnaire. Then, section 3.2.3 discusses prominent literature on ethical governance of AI to understand what other measures are necessary to effectively govern AI. Lastly, section 3.2.4 aims to create a better understanding of ethical codes of conduct and their effectiveness in changing behavior. This knowledge is important to understand the code of ethics from NLdigital, and eventually analyze its shortcomings.

3.2.1 Ethics theory

Generally, ethics can be seen as the study of moral principles. Traditionally, ethics researchers and philosophers have focused on the question of what the right moral principles are, and what is morally right and wrong (Roeser, 2018). These are complex questions because ethics are highly ambiguous. Still, influential researchers and philosophers have created different ethical theories over time that function as important theoretical frameworks for improving ethical decision making. They provide a guideline for argumentation in a field in which moral judgments of decisions and actions are normative, so that moral justifications can be made (Van de Poel & Royakkers, 2011). Ethical challenges with AI are widely observed but the way in which they are judged by individual companies in the ICT-industry remains elusive. The challenge and its impact are perceived differently as well as the 'correct' way of approaching them. To really understand AI ethics, it is, important to first understand the underlying ethical foundations. This way the nature of and solutions for ethical challenges in businesses can be understood, as well as the tension between different norms and values that may be conflicting. This section discusses these ethical foundations.

3.2.1.1 Normative ethics

This section mainly uses the chapter on normative ethics from Van de Poel and Royakkers (2011) and the Stanford Encyclopedia of Philosophy as a guideline to display the three most prominent ethical theories in normative ethics. Normative ethics is the study of how to act morally (Kagan, 2018). Normative ethics deals with the question of 'what ought to be' or how one ought to act rather than the question of 'what is'. Throughout history, philosophers and ethicists have introduced different theories of how one ought to act and live. The section below lays out the three most widely accepted and best-known theories.

- **Consequentialism:** in consequentialism, the consequences of an action are the only relevant principle in determining the morality of that action. Utilitarianism is the most popular type of consequentialism in philosophy. This theory that Jeremy Bentham developed, states that moral actions should bring the most happiness to the largest amount of people. A reason to use this approach is that it advocates happiness and has universal aspirations. Limitations are that the consequences of an action are largely unpredictable and that it poses problems for distributive justice (Sinnott-Armstrong, 2019; Van de Poel & Royakkers, 2011, pp. 78-89).
- **Duty ethics:** this theory of ethics does not base the moral judgment of an action on the consequences, instead an action should be in agreement with a moral rule, which can be a law, norm, or principle. These moral rules can come from religion or social

agreements such as a code of conduct. Duty ethics, or deontological ethics, were made noteworthy by Immanuel Kant in the 18th century. In Kant's specific view on duty ethics, he promotes autonomy and argues that through rational reasoning people are capable to decide the morality of an action themselves. Even though following universal ethical laws and moral rules can bind people to their duties, such as respecting others, this theory also has some limitations. First, Van de Poel and Royakkers (2011) use the case of a whistleblower to demonstrate that there can be conflicting norms. Second, they describe that deontological ethics can cause negligence to the consequences of actions (Alexander & Moore, 2007; Van de Poel & Royakkers, 2011, pp. 89-95).

- **Virtue ethics:** this field of ethics was established by Aristotle and puts 'good' characteristics and personality traits central in the morality judgment. The theory of virtue ethics prescribes that people can learn to be moral as they develop good characteristics such as courage and honesty. Because these traits can be developed, experiences and continued education play an important role from the perspective of virtue ethics. Van de Poel and Royakkers (2011) offer a set of virtues for engineers to be morally responsible. A limitation of this ethical theory is that it is not concrete enough for solving cases because it does not provide a method for judging the morality of individual action (Van de Poel & Royakkers, 2011, pp. 95-101).

These theories can be used as the universal viewpoint to judge the morality of actions, as is the case in moral monism or absolutism. For the descriptive part of this research in the ICT-industry, a more useful perspective is that of moral pluralism. Moral pluralism can be approached from two levels. First, moral value pluralism entails that there are many different moral values and context-dependent objective ethically relevant aspects (Mason, 2006). This pluralism addresses the objection of conflicting norms that Van de Poel and Royakkers (2011) described in Kant's deontological ethics. An influential pluralist within duty ethics was Ross, who believes in the plurality of prima facie duties. These duties are fidelity, reparation, gratitude, justice, beneficence, self-improvement, and non-maleficence (Ross, 2002). A second way of looking at moral pluralism is from that of a plurality of viewpoints. In this sense, one acknowledges that there are many different viewpoints that can be contradictory, but each worthy of consideration. Unlike the critique of monists that ethical pluralism follows the same philosophy of ethical realism, we support Mason (2006) in rejecting the view that there can be no right answer as is the case in moral relativism. Thus, by no means does this research subscribe to moral relativism and argue that this is a useful approach for AI ethics, but due to its descriptive nature, it does acknowledge that different relevant moral principles and viewpoints exist for ethical challenges of artificial intelligence. The scope of this research should, therefore, allow the opportunity to make these conflicting norms, values, and viewpoints visible. In the pursuit of a better code of ethics, these different viewpoints will finally be discussed and laid against widely accepted literature and practices in an ethical reflection to arrive to a first suggestion of moral principles for the next code of ethics.

3.2.2 Ethics of artificial intelligence

Over the last decade, many important books have been written about artificial intelligence, of which many have included a chapter dedicated to the ethical aspects of AI (Bostrom & Yudkowsky, 2014, pp. 316-334; Floridi, 2010; Russell & Norvig, 2016, pp. 1034-1040; Smith & Browne, 2019, pp. 191-211; Yao et al., 2018). This section will review some of the most prominent

works and views of AI ethics. This section ultimately aims to present and motivate the main concepts that this research focusses on and show the position of this research with regards to these concepts and issues. The scope for this review is mainly focused on AI in enterprises and is limited to narrow AI, it will thus not consider the ethics of ‘Artificial general intelligence’ and more philosophical topics such as superintelligence.

3.2.2.1 Framework

A useful ethics framework for AI comes from Dignum (2018). She splits ethical considerations for AI in three levels. First, ‘Ethics by Design’ deals with the integration of technical capabilities that allow artificially intelligent systems to ethically reason. Second, ‘Ethics in Design’ constitutes “the regulatory and engineering methods that support the analysis and evaluation of the ethical implications of AI systems as these integrate or replace traditional social structures”. Finally, “Ethics for Design constitutes the codes of conduct, standards and certification processes that ensure the integrity of developers and users as they research, design, construct, employ and manage artificial intelligent systems” (Dignum, 2018, p. 2).

Ethics by design lies out of the scope of this research because the code of ethics is aimed at guiding employee behavior within businesses and does not offer guidance for the integration of technical capabilities. Rather, this research focusses on a combination of ethics in design and ethics for design. Foremost, this research falls within ethics for design because it aims to provide valuable descriptive information for the next iteration of a code of ethics. This is a tool to promote integrity in the design and application of the AI system. However, the descriptive analysis also partly considers ethics in design as it analyzes the governance structures and development processes within businesses and their alignment with ethical principles.

3.2.2.2 Ethical challenges

The progress of artificial intelligence has caused the emergence of ethical challenges. Moreover, the increased adoption of AI technologies for applications in businesses and society has highlighted the impact these challenges can have, and thus the importance of solving them. Numerous academic articles (Bostrom & Yudkowsky, 2014; Jobin, Ienca, & Vayena, 2019; Lee & Park, 2018; Leikas, Koivisto, & Gotcheva, 2019; Stahl & Wright, 2018), books (Russell & Norvig, 2016; Smith & Browne, 2019; Yao et al., 2018), companies (Google, 2019; IBM, 2018; Microsoft, 2019) and institutions (HLEG AI, 2019; IEEE, 2016) describe ethical challenges with artificial intelligence. For example, Baylé (2019) discusses the most important ethical challenges for businesses in the design of AI systems in her series ‘Designing responsibly with AI’. In a concise but thorough analysis of AI ethics literature, Baylé offers fairness, transparency, human-machine collaboration, trust, accountability and morality as the most important ethical themes for responsible AI design (Baylé, 2019). Table 5 provides an overview of ethical challenges that researchers and experts have identified to be significant challenges for artificial intelligence.

Table 5: Ethical challenges of artificial intelligence emergent in different studies

Literature	Ethical challenges
(Baylé, 2019)	Fairness
	Transparency
	Job disruptions
	Trust

	Accountability
	Morality
(Bostrom & Yudkowsky, 2014)	Transparency
	Predictability
	Robustness
	Responsibility and accountability
	Auditability
	Incorruptibility
(Lee & Park, 2018)	Privacy
	Safety
	Human abuse
(Russell & Norvig, 2016)	Accountability
	Malicious AI
	Job disruptions
(Smith & Browne, 2019)	Fairness
	Reliability and safety
	Privacy and security
	Inclusion
	Transparency
	Accountability
(Yao et al., 2018)	Fairness
	Malicious AI
	Equal education
	Job disruptions
(Jobin et al., 2019)	Responsibility
	Safety and non-maleficence
	Transparency
	Fairness
	Privacy

As can be observed in Table 5, there is a clear convergence in literature towards a few key ethical challenges, namely fairness, trust and responsibility, privacy, safety and non-maleficence, and transparency. Our study, therefore, hypothesizes that these challenges will also play an important role for the members of NLdigital and thus for the companies in the Dutch ICT-sector. The interviews that will be conducted in this research will be the first method to check if this is indeed the case. The sections below will discuss each of these ethical challenges.

3.2.2.3 *Fairness*

Fairness is an ethical challenge that many businesses already have to consider in their decision-making. It is considered the equal treatment of different groups without favoritism or discrimination ("Fairness," 2019). Fairness is widely considered as one of the most important ethical principles in western societies. The right to equality and non-discrimination forms the first article of the Dutch constitution, as well as a fundamental element of international human rights law such as the Universal Declaration of Human Rights (UNG Assembly, 1948). Moreover, fairness (referenced as justice) can be found as one of Ross's prima facie duties (Ross, 2002).

For artificial intelligence specifically, fairness offers some novel challenges. Particularly the notion of algorithmic fairness in the context of artificial intelligence is extensively discussed in academia

(Altman, Wood, & Vayena, 2018; Chouldechova & Roth, 2018). According to Zheng, Dave, Mishra, and Kumar (2018) algorithmic fairness can be described as a constrained optimization that has the objectives of improving the efficiency and equity of decisions. Essentially algorithmic fairness aims for statistical parity for some variables across different groups. Algorithmic fairness is an important topic to discuss in the field of artificial intelligence because algorithms may act unfairly towards minority groups without any form of human intervention or control. Algorithms can, for example, treat groups differently with respect to gender, race, ethnicity, and disabilities (Zheng et al., 2018). Last year, for example, Amazon had to shut down its AI-powered recruiting tool because of latent bias. The machine learning algorithm that they used was trained with resumes from predominantly men and as a result taught itself that men were better candidates for Amazon's open positions than women (Dastin, 2018). Their algorithm was thus systematically favoring men over women. Other cases of systematic 'discrimination' by AI algorithms that are often talked about are credit allowance cases. Recently, for example, Apple and Goldman Sachs were the center of such a debate. Various individuals, including Apple co-founder Steve Wozniak, noted that the algorithm that the Apple Card uses discriminated against certain individuals. Regulators have now opened an investigation case into the credit card practices of the companies (Nasiripour & Natarajan, 2019). It becomes clear that building 'unfair' AI-systems can not only lead to unfavorable results for individuals and society, but also for the businesses that build them. It is thus in everybody's interest to build fair systems.

An important benefit of AI and machine learning for decision making in businesses is that decisions could be made without (human) biases. However, currently, biases are the main cause of algorithmic unfairness. Apart from biases, Chouldechova and Roth (2018) describe two other causes of unfairness in algorithmic fairness:

1. The minimization of the average error to fit the majority populations leads to a higher distribution of errors in the minority population
2. The dependency on actions taken in the past (particularly important in drug trials and recidivism prediction)

The way in which biased data causes biased recommendations or decisions by AI systems is because machine learning systems are trained with the biased data, and will reproduce or reinforce these biases (Chouldechova & Roth, 2018). AI giant Google describes three types of possible biases in AI systems, i.e. interaction bias, latent bias and selection bias (Google, 2018). The company has recently created a tool called 'What-if' to test for biases in your data so that they can be removed or minimized to increase the fairness of algorithms. Many researchers agree on the fact that collaboration and diverse teams are essential to minimize the number of biases in systems (Sanders et al., 2018; Yao et al., 2018, p. 46). Altman et al. (2018) propose a harm-reduction approach that aims to promote algorithmic fairness. This approach consists of an evaluation of the distribution of harm, a counterfactual analysis to explain and predict the consequences of algorithmic decisions. Google's 'What-if' can be seen as such a tool to minimize harm.

3.2.2.4 Responsibility, accountability, and trust

An important concept in ethics is that of responsibility. Particularly, in the case of contemporary technologies such as AI and autonomous artifacts, responsibility and accountability have become

increasingly ambiguous. Van de Poel and Royakkers (2011, p. 10) defined moral responsibility as “responsibility that is based on moral obligations, moral norms or moral duties”. Miller (2011) offers a working definition of moral responsibility distinctively for computing artifacts. In this definition, people are “obliged to adhere to reasonable standards of behavior”. Miller also presents that creators or users of these artifacts have to respect those who could be affected by the technology and should be answerable for their actions (Miller, 2011). The tables below delineate different forms of responsibility that are identified in the literature to provide an understanding of the range of different forms of responsibility that might be encountered during research on this ethical aspect in the industry. Table 6 provides an overview of some of the different forms of responsibility that can be discussed in ethics (Van de Poel & Royakkers, 2011, pp. 7-18).

Table 6: Overview of different responsibility concepts

Responsibility	Description
Professional	The duty that one has in his role as a professional to execute his or her job within the limit of what is morally tolerated, e.g. without doing any harm.
Role	Often informal rules and agreements based on a role one plays in a certain context, e.g. the responsibility one has towards his or her family.
Active	Responsibility of an individual or collective before something undesirable has happened. Negative consequences of actions can be avoided by the way people act.
Passive	Responsibility of an individual or collective after something happened. This may include types such as accountability, blameworthiness, and liability.

Van de Poel and Royakkers (2011, pp. 10-13) separate three different forms of passive responsibility. An overview of the different types of passive responsibility is given in Table 7.

Table 7: Types of passive responsibility

Type	Description
Accountability	Show accountability and provide justification for one’s actions and the subsequent consequences after an undesirable event.
Blameworthiness	Someone can legitimately be blamed for one’s actions. This can be established by checking these conditions: wrong-doing, causal contribution, foreseeability, and freedom.
Liability	The responsibility that one has to take as dictated by the law. People that are considered liable for their actions are generally penalized.

Notable is that in the context of artificial intelligence researchers tend to often focus on accountability instead of responsibility and the words are sometimes used interchangeably.

Matthias (2004) observes significant challenges with responsibility for AI. He argues that a “responsibility gap” exists for self-learning algorithms (Matthias, 2004, p. 177). Algorithms are not completely coded and defined by engineers anymore, but instead, learn from new data. Because no one has absolute control over the decision that a learning algorithm makes, the traditional

model for algorithm responsibility in which systems are well-defined and predictable does not hold anymore. This can result in the fact that nobody will assume responsibility for those decisions (Matthias, 2004; Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016, p. 11). The responsibility gap can form a problem for active responsibility as well as the three different forms of passive responsibility. The 'black box' dilemma is an important contributor to this responsibility gap and is an ethical aspect often mentioned in research with regards to AI (Baylé, 2019; Manyika & Bughin, 2018). The vagueness and complexity of self-learning algorithms make accountability and responsibility a serious problem (Bostrom & Yudkowsky, 2014, p. 319). Moreover, people do not trust algorithms because they are viewed as a black box and people cannot understand them.

One notion to avoid such a responsibility gap is that of 'meaningful human control'. Santoni de Sio and van den hoven (2018) argue that for highly autonomous systems such as AI-systems, human control, and human responsibility will be necessary to "avoid unreasonable risks and provide a place to turn to in case of untoward outcomes" (Santoni de Sio & van den hoven, 2018, p. 2). In their work, they identify two conditions for such control, i.e., tracking and tracing. The first refers to the system being capable of replying to moral concerns, whereas the latter describes the need to be able to trace back the decisions, operations, and considerations of an outcome by at least one human.

Kroll et al. (2016) argue that white-box testing, an audit in which the auditor can assess both the inputs and outputs as well as the code, can be a solution to the limitations of accountability with AI (Kroll et al., 2016, pp. 23-26). Salvino (2018) also sees auditing as a key mechanism to explain decisions and accountability by machine learning systems. He claims that the ability to audit a system is an ethical requirement and a standard practice in business and should therefore also apply to artificial intelligence systems. Yao et al. (2018, pp. 197-201) argue that enterprises (even without an audit) need to be aware of their ethical responsibility and should constantly assess how their AI systems affect customers, employees and society.

3.2.2.5 *Privacy*

Another ethical challenge that has received much attention in recent years is that of privacy. In an analysis of privacy principles in the context of learning analytics in institutions, Pardo and Siemens (2014) identified four privacy principles that should be discussed in the deployment of analytics systems (e.g. machine learning systems): transparency, control over the data, security, and accountability and assessment. According to Pardo and Siemens (2014), transparency means that all stakeholders should be informed of the way data is collected, stored, transferred and processed, and what kind of analytics are done with the data. Control over data refers to "the right of users to access and correct the data obtained about them" (Pardo & Siemens, 2014, p. 446). Security refers to keeping the data safe and protecting it so that it cannot be breached and abused by other entities to which users have not given consent. Lastly, dealing with privacy entails to "identify entities that are accountable for specific data and analytics areas" (Pardo & Siemens, 2014, p. 447). Finally, Pardo and Siemens (2014) describe the responsibility of organizations to continuously evaluate and refine these four principles. Perhaps the most notorious example of infringement of these privacy principles was in the digital campaign of the 2016 Trump campaign. In the campaign, Cambridge Analytica, a company that was hired to help bring victory to Donald Trump, scraped data from over 50 million Facebook profiles to build psychological profiles to swing votes by targeting individuals with personalized political advertisements (Gadwalladr &

Graham-Harrison, 2018). First of all, we believe this violated the transparency principle by systematically not informing stakeholders of the collection and use of their data. Moreover, even if users were aware that their data was being used, they had no option of accessing the data let alone correct it.

Because the performance of machine learning algorithms is largely dependent on the quality and availability of large scale data sets, fundamental to unlocking the potential of artificial intelligence will be the way in which privacy sensitivities are handled so that scandals like these can be avoided (Montjoye, Farzanehfar, Hendrickx, & Rocher, 2017). Stahl and Wright (2018) suggest using the 'Responsible Research and Innovation' (RRI) approach in response to privacy challenges with artificial intelligence in information systems. Primarily, this entails engaging stakeholders from early on in the innovation process and being open to their values, needs, and thoughts. Stahl and Wright (2018) discuss different approaches to RRI and provide a practical example of how it can be applied. A technical solution for privacy comes from Montjoye et al. (2017). The authors observe how de-identification, a solution that was previously used as a remedy, does not scale to large data sets that are often required for sophisticated machine learning algorithms. Stahl and Wright (2018) argue that only by rethinking how we protect data, such as by using novel privacy-enhancing technologies such as 'Open Algorithms (OPAL)', we can reap the benefits of AI while protecting the privacy rights of individuals.

3.2.2.6 Safety and non-maleficence

Non-maleficence refers to the obligation to avoid or minimize harm (Stahl & Wright, 2018). Non-maleficence is not only important for AI ethics but is generally considered an important norm in society. It is one of the four 'prima facie' duties that were described by W.D. Ross (Ross, 2002). The difficulty for non-maleficence with AI arises because full autonomy and control are lost, and a responsibility gap exists. The notion is often discussed in the context of AI for medical applications (Anderson, Anderson, & Armen, 2006; Keskinbora, 2019). More general literature in the context of AI emphasizes the need for safety and security and the notion that AI should always avoid foreseeable or unintentional harm (Jobin et al., 2019). The results of the meta-analysis from Jobin et al. (2019) indicate that the majority of academic articles interpret harm as discrimination, violation of privacy or physical harm. Examples of maleficence of AI include a twitter bot that became verbally abusive, an adult content filtering system that failed to block inappropriate content, and deadly incidents with autonomous robots and self-driving cars (Yampolskiy & Spellchecker, 2016). Incorporating advanced security and cybersecurity techniques will help to prevent some of these safety breaches, but Yampolskiy and Spellchecker (2016) suggest that the frequency and impact of these incidents will increase over time and that systems can never be entirely safe. In this case, literature proposes that risks should be assessed, reduced, and mitigated and that the attribution of liability should be clearly defined (Jobin et al., 2019). The European Union high-level expert group on AI also states that safe systems should have a fallback plan in case of negative repercussions (HLEG AI, 2019).

3.2.2.7 Transparency and explainability

An important challenge that has developed simultaneously with the rise of more sophisticated artificial intelligence is a lack of trust in applications using artificial intelligence. Both end-users, as well as professionals working with AI, can develop distrust due to the lack of transparency in the decision-making process. The reasoning process is often a vague black box that provides no

insight for users and other stakeholders as to how decisions are made. This lack of transparency forms a major barrier to the widespread acceptance and adoption of AI by both professionals and society (Owotoki & Mayer-Lindenberg, 2007). Another reason why transparency is important is that it forms the groundwork for many of the other ethical challenges of AI in businesses. If an undesirable or harmful event occurs, transparency is needed so that the decision-making and circumstances can be understood and used as evidence of how something happened (Bryson & Winfield, 2017). In that respect, transparency is also needed to assess the responsibility and accountability of different actors. Furthermore, by proactively being transparent, people could see 'bad' decisions coming, and some accidents can be avoided altogether. Transparency is also an important concept for ensuring fairness of models. According to Olhede and Rodrigues (2017, p. 1) "we cannot evaluate the probity, or fairness, of a computer-generated decision without first having a clear understanding of the population from which the data is drawn and the logical steps an algorithm goes through to determine its outputs". Being transparent about these two points will increase the fairness of AI-systems and create a sense of trust with the general public.

However, the problem remains that currently much of what an algorithm does is invisible. This often causes AI-systems and their creators to be inscrutable to those whose data it feeds on (Olhede & Rodrigues, 2017). For the data part, the GDPR was created that officially regulates the consent that users should give and the level of transparency with regards to people's data that is required from companies. Most of the GDPR is focused on data rights and data processing, but another important sub-concept of transparency is briefly addressed; i.e., explainability. Article 22 of the GDPR describes that a data subject has the right to not be subject to the decision process when a decision is based only on automated decision making (EU, 2016, Article 22). Such systems that operate without any human intervention are almost always using some form of artificial intelligence. The difficulty with this article is that it does not specify which information should be provided and that the vast majority of contemporary AI-systems do operate with some degree of human intervention. This has caused the right to explanation in the GDPR to be heavily debated (EU, 2016, Article 15). Although adequate means to enforce explainability of AI-systems decisions and processes do not exist, it is still very important that such systems are explainable. Explainability is important for the same reasons that transparency is important but provides even an extra dimension because it also includes the ability of people to understand AI systems. Došilović, Brčić, and Hlupić (2018) view interpretability and explainability as the same concepts and split the concepts into two categories; integrated or transparency-based explainability and *post-hoc* explainability. Integrated signifies to bake in a degree of transparency into your systems, whereas *post-hoc* explainability means providing interpretable explanations for decisions that were made by 'black-box' AI-systems.

Concluding, designing and deploying ethical systems is one of the key challenges of artificial intelligence today. Providers of AI systems or products will need to take seriously all the challenges described in this section. The predominant challenges that were found in this literature study are fairness, responsibility, privacy, transparency, and safety/non-maleficence. We anticipate that at least some of these challenges will be considered as pressing challenges by business during the interviews and questionnaire. Although enterprises bear a large ethical responsibility to deal with those, national and international policies and control mechanisms may

be needed to bind actors to this responsibility and force ethical behavior. The next section (3.2.3) will discuss the legal safeguards and control mechanisms that are in place.

3.2.3 AI Ethics governance

3.2.3.1 *Legal safeguards, governance structures, and control mechanisms*

This sub-section maps out the legal landscape of artificial intelligence and discusses works on ethical governance of artificial intelligence within organizations. Ma, Zhang, and Zhang (2018) offer possible legal issues in different application domains of artificial intelligence, such as for autonomous driving, surgical robots, medical diagnosis, investment consultants, smart identification and more. However, not many laws regarding artificial intelligence and machine learning exist to provide clear norms for AI practices in such domains. Canada and the United States are examples of countries that have passed bills to put aspects of Artificial Intelligence into law. Nonetheless, these laws still focus mainly on autonomous driving and the protection of personal information (Soares, Ahmad, Levush, Guerra, & Martin, 2019). This means that ethical challenges such as fairness, explainability, responsibility, and perhaps also safety are not represented in regulation. This research will not further investigate the need for such regulation, but rather examine the need for additional ethical business tools and mechanisms.

In 2016, the European parliament passed a new European regulation called the General Data Protection Regulation (GDPR). The regulation is designed to protect the data and privacy of citizens of the European Union and the European Economic Area, both inside and outside of the European Union (EU, 2016). The GDPR makes it possible to fine companies that infringe the law with fines of up to 20 million euro's or 4% of their global revenue if this value exceeds the 20 million euro's bar. According to Goodman and Flaxman (2017), there are two parts of the GDPR that are tightly connected to machine learning. First, they describe the deeply embedded European right to nondiscrimination. This poses challenges for machine learning programs, because these are often trained with data from society, and if the data contains inequalities and forms of discrimination, these will also be present in the decision-making process. This may finally lead to systematically biased decisions and discrimination. Second, the right to explanation, that according to many researchers does not even exist, is at most vaguely addressed in articles 15 and 22 of the GDPR (EU, 2016). This right would be inherently challenging for artificial intelligence, and more specifically the newer more advanced deep learning algorithms. Goodman and Flaxman (2017) ask the reader in their study "what hope is there of explaining the weights learned in a multilayer neural net with a complex architecture?" (Goodman & Flaxman, 2017, p. 55). This question exemplifies just how complex it is to explain the black box processing that happens in advanced AI algorithms.

Gasser and Almeida (2017) offer a three-layered approach for thinking about AI governance (Figure 3). The model provides a guideline for what government proposals for policy and legislation should deal with. First, standards for data governance, algorithms and algorithm accountability should be created. The GDPR also falls within this first technical layer. Next, policymakers and overarching organizations can focus on aligning AI applications with ethical principles. The second layer will become increasingly important when maturing AI application will emerge. Lastly, specific bodies and authorities for AI governance are created to translate the principles from the first two layers into legal frameworks. An international committee as a curator

is finally proposed by Gasser and Almeida (2017) to establish and align global principles for AI. Gasser and Almeida (2017) offer the idea that overarching organizations will be helpful in finding principles and norms for AI. An interesting analogy can be drawn with NLdigital and the ICT industry. NLdigital could take the role of curator for the ICT-industry and establish and align industrywide principles for AI. The code of ethics that they introduced can be seen as such a formalized form of ethics principles for the industry. The framework of Gasser and Almeida (2017) thus underlines the importance of NLdigital's code of ethics for the industry.

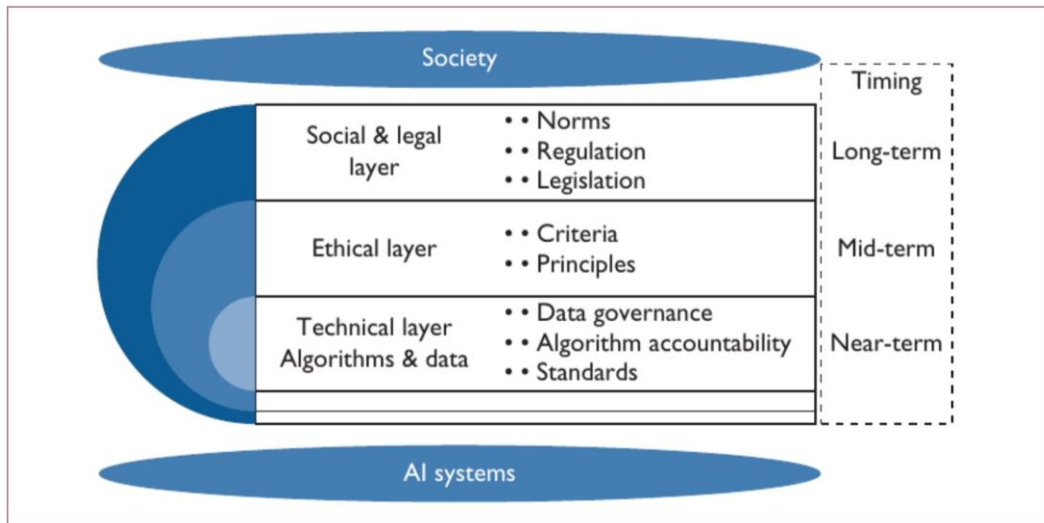


Figure 3: Layered model for AI governance, adapted from: (Gasser & Almeida, 2017)

3.2.3.2 Organizational governance of ethics

Apart from the need to think about AI governance from a legal and policymaking perspective, governance from the perspective of companies and collectives is also an important concept to discuss. Winfield and Jirotko (2018) propose five key pillars for solid ethical governance of AI in organizations. Organizations should according to their research:

- Publish a code of ethics, alongside a ‘whistleblower’ mechanism
- Provide ethics and responsible innovation training for everyone in the organization
- Practice responsible innovation, i.e. actively involve a wide range of stakeholders through a framework
- Be transparent about ethical governance
- Really value ethical governance, i.e. everyone in the organization should show this as one of their core values

These pillars provide valuable information for the interview and questionnaire of this research. It suggests that organizations should be working with at least one set of principles, guideline or codes of ethics and have a whistleblower mechanism in place. If companies do not work with a set of principles yet, NLdigital's code is a valuable tool to fill that void. With regards to a whistleblower mechanism, if companies do not have such a channel in place already, this should be demanded through the inclusion of a principle for whistleblowing in the code of ethics. Another option could be to create an industrywide independent whistleblowing vehicle. NLdigital is an organization that could set-up such a vehicle parallel to its code of ethics. The pillars also suggest

that ethically responsible companies are discussing ethical concerns with other stakeholders around them such as customers or suppliers. Lastly, according to these pillars, we can expect companies to be aware of the need for good ethical governance at all levels of the organization, e.g. in top-level management and in the roles of more practical daily decision-makers. Concluding, these pillars form an important checklist for companies in the ICT-industry for 'solid' ethical governance of AI. This framework will be used in both the interviews and survey to assess where companies are in terms of ethical governance. This study expects large multinational corporations to possess more pillars than small and medium-sized companies with a smaller footprint. The results of the interviews and survey will inform the need for additional tools and mechanisms to ensure good ethical governance of AI within these different sizes of ICT-organizations.

3.2.4 Codes of Ethics

In this section, an analysis is given of codes of ethics as a theoretical domain. First, the concept of an ethical code is defined in section 3.2.4.1 In the analysis, the effectiveness and drivers of an ethical code of conduct (section 3.2.4.2), factors that can inhibit and enhance the adoption of codes (section 3.2.4.3), and literature on ethical codes specifically for artificial intelligence (section 3.2.4.4) are delineated. Finally, section 3.2.4.5 discusses the importance of whistleblowing in the case of ethical misconduct.

3.2.4.1 *Definition*

Stevens (1994) provides a descriptive definition of a code of ethics by describing it as a written document intended to influence employee behavior, which can range from one paragraph to fifty pages. By combining elements from Steven's and several other definitions Schwartz (2004, p. 324) suggests the following working definition: "a code of ethics is a written, distinct, and formal document which consists of moral standards used to guide employee and/ or corporate behavior". Throughout the rest of this research, this definition is adopted for codes of ethics.

3.2.4.2 *Drivers of an ethics code for organizations*

Myriad researchers in the last two decades have tried to establish the effectiveness of codes of conduct (Adam & Rachman-Moore, 2004; Erwin, 2011; Paine, 2004). A large body of the scientific literature agrees that if ethical codes are designed and implemented adequately, the code may have a significant impact on the ethical behavior of people in an organization (Adam & Rachman-Moore, 2004). Thereby it becomes an essential factor in the success of business corporations (Paine, 2004). Apart from the value that influencing ethical behavior can have on a business, Adam and Rachman-Moore (2004) give other reasons for why businesses may adopt ethical codes. One of the benefits of an ethical code is that it can help to overcome the different perceptions and (cultural) norms of people within an organization. Some companies use ethical codes because it is standard business practice in the industry, or because it is mandated by regulations. Lastly, companies use it to enhance their public perception and appearance. Just like Adam and Rachman-Moore (2004), Erwin (2011) also supports that good and adequately designed codes of conduct are an effective tool to influence ethical behavior. He proved this by analyzing the relationship between companies with a 'good' code of conduct and their corporate social responsibility (CSR) ranking. Erwin used 8 components (i.e., public availability, tone from the top, readability and tone, non-retaliation and reporting, commitment and values, risk topics, comprehension aids, and presentation and style) that determine the quality of a code to compare

how the companies with these codes performed in rankings of corporate responsibility, e.g. ethical behavior and public perception. Erwin concludes that companies with high quality codes of conduct ranked significantly higher in CSR and ethical rankings, and thus that rather than just having a code of conduct, code quality plays a key role (Erwin, 2011). For the purpose of this work, the literature above has important implications. First, it supports the notion that the quality of NLdigital's code of ethics will influence its ability to influence behavior, and thus its adoption by organizations. Erwin's quality components can be used for recommendations for the next iteration of the code of ethics. Especially, the two heaviest weighing (20% each) components which are readability and tone, and the addressing of all relevant risk topics, will be key topics to assess during the interviews.

3.2.4.3 Adoption of ethical codes within organizations

A large body of academic research on ethical codes focusses on the adoption of ethical codes by an organization's employees (Adam & Rachman-Moore, 2004; Stevens, 2008; Webley & Werner, 2008). Stevens (2008) argues that ethical codes can be a very effective strategic management tool in stimulating ethical behavior within companies. He observes that the success of the code is highly dependent on two key components:

- the code should be embedded in the organizational culture.
- the code should be communicated effectively between the different layers of the organization.

However, in practice, companies do not always do those things successfully. Webley and Werner (2008), for example, observe a clear gap between ethical principles and ethics practice in organizations. According to Webley and Werner, the reasons for this are deficiencies in organizational culture and ineffective ethics programs. The latter is made up of inadequate codes of ethics and a lack of communication and embedding of the code in the different layers of the organization. For this research, inadequate code design is an important concept. It suggests that the level of ethical behavior in AI practices can be increased by adequate code design and better aligning the content of the code with practical challenges from the industry. Webley and Werner (2008) give three examples of poor code design; 1) they do not encompass all the duties to the stakeholders in the network, 2) the code consists only of a practical compliance-based set of rules but does not provide guidance 3) the code does not cover the key decision-makers of a company (Farrell & Cobbin, 1996). Instead, Webley and Werner (2008) underline the significance of having a value-based code that includes both the obligations of the employees as well as the management. Like Stevens, they also stress the importance of deeply incorporating an ethical code in the organizational culture and propose that organizations should create formal ethics programs such as "ethics training, mechanisms to seek ethics advice, and means to report misconduct anonymously" (Webley & Werner, 2008, p. 408). Adam and Rachman-Moore (2004) argue the contrary in their research about employee attitude towards the implementation of ethical codes. Their results indicate that informal methods are more effective in gaining employee commitment to ethical codes of conduct. Rather than only formal training, they stress the importance of committing to the code through social interactions such as a manager that sets the right example (Adam & Rachman-Moore, 2004).

For this work, the abovementioned literature suggests that there are ways in which the gap between NLdigital's ethics principles and ethics practice in the industry can be reduced. On top of making a more adequately designed code that our study is largely focused on, formal ethics programs such as ethics training, anonymous channels for misconduct, and feedback channels are necessary to reduce this gap.

3.2.4.4 Codes of conduct for AI

Progressively more companies are adopting ethical codes specifically for the use of artificial intelligence. However, specific research on codes for artificial intelligence is still lagging behind. Few studies on the effect and quality of ethical codes for the use of artificial intelligence have been conducted. Its effectiveness and industry-specific drivers for adoption are therefore hard to establish. Still, the most complete work on AI ethical codes of conduct comes from Boddington (2017). In her book, she analyzes experiences with ethical codes from the past to infer useful insights with regard to ethics codes for AI. She finally provides a set of recommendations for how to proceed with the development of AI-specific codes of conduct. Boddington (2017) proposes 5 key procedures for the creation and implementation of ethics codes:

- Diversity in participation throughout the life cycle of the development of an ethics code.
- Transparency of the process and people with important ethical responsibilities.
- Excellent communication with internal and external stakeholders.
- Mechanisms for revision and critique such as whistleblowing structures.
- The timing of the introduction of codes should not be hurried.

The procedures that Boddington offers support the necessity for this research. Namely, the first procedure suggests that the participation of high-level executives and experts from a select group of large corporations and institutions is not diverse enough to develop a code of ethics. A wider range of companies (e.g. small and medium-sized enterprises), experts and job levels of employees should be included in the development of the ethics code. The interviews and survey that are conducted for this research will, therefore, put a heavy emphasis on gathering data from small and medium-sized enterprises (SMEs). Like Winfield and Jirotko (2018), Boddington (2017) also presents the inclusion of (anonymous) mechanisms for critique such as a whistleblowing structure as a key procedure. The repeated suggestion of such mechanisms in the context of AI, as well as in a wider context, makes this an important topic for investigation for this research. Section 3.2.4.5 discusses whistleblowing in more detail. The other procedures that Boddington (2017) proposes fall out of the scope of this research. Good communication is considered a shared duty by the designer of the code of ethics and the company but is not influenced by the design of the code, timing is not a variable anymore as the code was already introduced, and transparency of the process will be ensured by the proper documentation of this research.

Boddington also provides a set of suggestions with regard to the content of ethical codes. These include the inclusion of industry/application-specific issues, responsibility and accountability structures, legal frameworks, and specific actionable items (Boddington, 2017). This last notion contrasts with the view of Webley and Werner (2008) who argued that value-based codes of ethics are more effective than compliance-based codes. Boddington (2017) recognizes the inherent tension between generality and specificity of ethics codes but believes that in the specific

case of AI different levels of specification are a necessary condition. Actionable and practical steps should guide engineers in the development of AI (Boddington, 2017).

Finally, Boddington warns for the limitations of codes of ethics for AI. First, a formal code for organizations can be dangerous as it may put companies on a defensive stance. Companies may want to avoid complications, publicity, and control. Some organizations will prefer to stay out of trouble than be obsessed with complying with the code. Another limitation that Boddington (2017) describes is that some organizations will work up to the code. Their ethics practices will be limited to complying with a code, but they may not go beyond the code even if ethical challenges behind such codes will arise. Boddington (2017) calls this an “ethical tick-box culture”. Lastly, organizations that see and express themselves as very moral actors (e.g., organizations that adopt a multitude of codes of ethics) may actually decrease the likelihood of ethical behavior. Such actors may “overinflate their self-assessed moral character” (Boddington, 2017).

3.2.4.5 Whistleblowing for neglecting ethics in artificial intelligence

The importance of channels for reporting misconduct has come forward in different publications discussed before. In section 3.2.3.2 the five key pillars for solid ethical governance of AI in organizations by Winfield and Jirotko (2018) were discussed. One of these pillars was the creation of a whistleblower mechanism alongside a code of ethics. Furthermore, in section 3.2.4.4 Boddington (2017) proposed a mechanism for revision and critique such as whistleblowing structures as a key procedure for the creation and implementation of ethics codes.

According to Bowden (2008), a whistleblowing mechanism is an important method to raise attention to breaches of a code of ethics. It gives employees of an organization a way to report misconduct anonymously and without fear of retaliation. Whistleblowing channels can be hosted by different means and organizations. Organizations can have channels internally, such as a confidential person, or in the form of an ICT-tool that can enable anonymity while filing for misconduct (Hussien & Yamanaka, 2017). Smaller organizations typically do not have such channels, whistleblowers could then go to the media, government, or law enforcement. However, this can harm the company significantly. Bowden (2008), therefore, proposes that professional and trade associations could act as whistleblower support centers for their respective industries. Trade associations are in a strong position to oversee and ensure ethical behavior because of the credibility they provide to their member companies. Although limited in their capabilities to provide ‘tough’ sanctions such as fines, they could still take up penalties to the extent of retraction of a company’s membership. However, professional and trade associations are often partially or fully funded by their members and can thus have a strong aversion of imposing membership withdrawals. Still, according to Bowden (2008), “industry associations and professional societies are the strongest arbiters of behavior in an industry, and it behooves them to apply sanctions that commensurate with this responsibility”. The arguments provided by Bowden (2008) provide an interesting case for NLdigital to pick up such a role for the Dutch ICT-industry, at least if a substantial part of the sector does not have access to such whistleblowing channels yet.

Sub-research question 1: What are the chief ethical aspects in the research, development, and application of AI-systems for businesses in the Netherlands that are identified in the specialized literature?

Answer: Important ethical aspects in the research, development, and application of AI-systems can be divided in concrete ethical challenges that businesses have to deal with, prevent, or be aware of on the one side, and governance of these ethical challenges on the other side. The most prevalent ethical challenges that are identified are:

- Fairness
- Responsibility, accountability, and trust
- Privacy
- Safety and non-maleficence
- Transparency and explainability

In terms of governance of these ethical challenges with AI, the academic literature offers different aspects for solid ethical governance (Winfield & Jirotko, 2018). First, businesses are expected to publish a code of ethics, alongside a 'whistleblower' mechanism. This underlines the significance of NLdigital's code of ethics, and the importance of our work in improving this code, especially for small and medium sized enterprises. Other important ethical aspects for good governance of AI include the availability of a whistleblower structure, ethics and responsible innovation training for everyone in the organization, and transparency about the form of ethical governance. The interviews and questionnaire will determine to what extent these aspects are present within companies in the industry.

3.3 Ethical guideline delineation

This section analyzes the key principles and challenges among a variety of codes of ethics for artificial intelligence. The goal is to identify shortcomings and room for improvement in NLdigital's code of ethics. Ethical challenges addressed in other codes, the inclusion of principles, and specificity of codes (i.e., value-based or compliance-based) will be analyzed to identify areas that require further empirical analysis in the interviews and questionnaire. In other words, this section informs the data that will be gathered in the second part of the qualitative phase and the quantitative phase by searching for both common and novel ethical building blocks between NLdigital's code and other codes of ethics. First, research from Jobin et al. (2019) that analyzes common elements among 84 ethics principles and guidelines is discussed in section 3.3.1. Next, a deep dive is done into different codes of ethics and ethics guidelines. The code of ethics from NLdigital is thoroughly analyzed in section 3.3.2, the European guideline for trustworthy AI in section 3.3.3, and ethics principles from a set of companies in the industry are briefly examined in section 3.3.4 (Google, 2019; IBM, 2018; Microsoft, 2019). Finally, these will all contribute to the creation of a conceptual overview of ethical challenges with AI in section 3.4.

3.3.1 The landscape of AI principles and guidelines

Recent research by Jobin et al. (2019) has drawn a global landscape of ethics guidelines for Artificial Intelligence. Their goal was to study whether an agreement on certain principles exists by analyzing ethical principles and guidelines from different organizations. Looking at this study is valuable for our research because it quickly provides an extensive analysis of a large part of the AI ethics principles, guidelines, and codes landscape. The result of the analysis of 84 principles and guidelines is an extensive table of ethical principles (Table 8), which this research calls ethical building blocks. From the frequency that these blocks were represented in different ethics guidelines, Jobin et al. (2019) observe convergence to five key principles, i.e. transparency, justice/fairness, non-maleficence, responsibility, and privacy. However, they also conclude that there is a strong discrepancy in how these principles are understood. For our research, these principles provide an important benchmark of which principles we can expect to be present in NLdigital's code of ethics. If these principles are not present, it either means that they were forgotten and that the code can thus be improved, or that not all of these principles are important or applicable to the ICT-industry. The hypothesis of this work is that the code can be improved. This will be tested by proving that they are indeed relevant challenges in the Dutch ICT-industry.

Based on the findings in their research, Jobin et al. (2019) finally recommend that the development of guidelines should be based on extensive ethical analysis and should include relevant implementation strategies. The implication of these recommendations for our study is twofold. First, it supports the driver of this work that more extensive ethical analysis is necessary to improve the current code of ethics code for the ICT-industry. Especially, a more thorough analysis of the ethical challenges of SMEs is necessary as these were largely overlooked in the creation of the first edition of the code. Second, it offers the importance of implementation strategies for a code of ethics and the importance of mechanisms that communicate, encourage, and enforce ethical behavior within organizations. The current strategies, mechanisms, and tools in the ICT-industry will be enquired in the interviews, with the goal to identify gaps and room for improvement.

Table 8: Analysis of 84 ethical principles and guidelines, adapted from: (Jobin et al., 2019)

Ethical principle	Number of documents	Included codes
Transparency	73/84	Transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing
Justice & fairness	68/84	Justice, fairness, consistency, inclusion, equality, equity, (non-)bias, (non-)discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution
Non-maleficence	60/84	Non-maleficence, security, safety, harm, protection, precaution, prevention, integrity (bodily or mental), non-subversion
Responsibility	60/84	Responsibility, accountability, liability, acting with integrity
Privacy	47/84	Privacy, personal or private information
Beneficence	41/84	Benefits, beneficence, well-being, peace, social good, common good
Freedom & autonomy	34/84	Freedom, autonomy, consent, choice, self-determination, liberty, empowerment
Trust	28/84	Trust
Sustainability	14/84	Sustainability, environment (nature), energy, resources (energy)
Dignity	13/84	Dignity
Solidarity	6/84	Solidarity, social security, cohesion

3.3.2 Analysis of NLdigital's ethical code

This section provides background information about NLdigital's code of ethics, as well as an analysis of the principles in the code. The ethical code of conduct was created by NLdigital, whose official name was still 'Nederland ICT' at the time of creation. The ethical code was a result of cooperation over the course of around a year with 8 members companies of the trade association; Betabit Nederland BV, Centric Netherlands BV, CGI Nederland BV, Dell BV, Facebook Netherlands BV, Google Netherlands BV, Microsoft BV, and Ordina Nederland BV. These companies, predominantly large multinationals, were all part of the 'think-tank ethics' of NLdigital. Four companies, in particular, were actively part of discussions and dialogues in the creation of the code. Experts in these discussions included executives from Microsoft, IBM, Centric and Betabit. Before the code was made public, NLdigital aligned it with the revision of the 'Ethics Guidelines for Trustworthy AI' by the European Commission (EC) that was entering its pilot phase a week later (HLEG AI, 2019). This guideline was developed from the doctrine that artificial

intelligence should be lawful, ethical and robust. In the document, the EC presents the seven basic requirements that trustworthy AI should meet (section 3.3.3).

Afterward, NLdigital consulted with the digital ethical council for their opinion on the ethical code. This council is an external advisory council of the trade association consisting of experts in ethics and digitalization from different institutions (TU Delft, TU Twente, Leiden University, and Clingendael Institute). After a series of quick adjustments, and a sign off by NLdigital's board of directors, ultimately, the 'Ethical Code for Artificial Intelligence' was handed over on a mirror by Peter Zijlema (managing director of IBM Netherlands) on the 21st of March to State Secretary for Economic Affairs and Climate Policy Mona Keijzer.

The Ethical code is a non-binding code and was designed specifically but not solely for its members. The goal was to make the guideline broadly applicable. For this reason, NLdigital has deliberately used a broad definition of artificial intelligence in the creation of the code, i.e. everything that touches self-learning or self-reasoning systems (van der Beek, 2019). The code has some flexibility build in to be able to adapt to societal and technological developments. According to NLdigital, the guideline is a living document that may be reviewed "as the need arises" (NLdigital, 2019). The next section (3.3.2.1) describes the type of organizations that are in NLdigital's network. These are essentially the businesses for which the code was created. Then, section 3.3.2.2 analyzes and displays the eight principles that the code consists of and the overarching ethical building blocks.

3.3.2.1 Type of organizations within NLdigital's member network

Although NLdigital's members are mostly recognized as being part of the ICT-sector, most of them have different characteristics and offer different products and services. First of all, the members of the trade association differ in size. Member companies include large national organizations, large multinational organizations, medium-sized national and international organizations, as well as small organizations and start-ups.

Another important way in which we can distinguish the member organizations of NLdigital is by the types of products and services that they offer. As the abbreviation ICT suggests, organizations are working mostly with information and communication technologies. The vast majority offers an IT service to its clients. This can be by providing service as a consultant, developing software tools, and managing software and IT systems, providing cloud services, etc. Another important area where members are active is in the communications sector. Businesses active here typically create chatbots, provide mobile networks, provide communication networks, provide internet connections, develop 5G networks, or create electronic communication devices such as mobile phones, routers, computers, etc.

3.3.2.2 Ethical building blocks in the code of ethics

This section analyzes the overarching ethical building blocks in the code of ethics. Table 9 displays the principles of the code, the building blocks and the basis (i.e. value-based or compliance-based) of the code as defined by Webley and Werner (2008). The eight principles are taken verbatim from NLdigital's code of ethics. The ethical building blocks and classification of the base are interpretations of this research of the important ethical concepts that underly the eight principles. Thus, these are not part of the original code of ethics documents. Because the

code was created for the trade association its members, all principles start with the addressing of “the member company” (NLdigital, 2019).

Table 9: NLdigital’s code of ethics and ethical building blocks

No.	Principles	Ethical building blocks	Base
1	The member company is aware of and anticipates the influence that implementation of AI can have on public values, such as honesty, equity, and inclusivity, as well as the principles of ease of explanation and use of (and rights to) the data.	Awareness, active responsibility, foreseeability, explainability, privacy	Value
2	The member company makes it clear within the chain when a user is dealing with AI, as well as the responsibility that each party in the chain bears when applying AI.	Responsibility, transparency	Compliance
3	The member company is aware of and transparent regarding the state of the technique for AI and its technical possibilities but is also aware of the technical limitations of applications and the necessity to communicate clearly regarding this. This leads (among other things) to minimizing undesired "bias" and to promoting inclusive representation.	Awareness, 'active' transparency, fairness	Value
4	The member company provides insight into the data that is used by the AI application or offers the possibility for the other chain partners to acquire such insight.	Transparency, privacy	Compliance
5	The member company provides the (technical) possibility for all parties who are directly involved to trace the recommendations or decisions made by its AI systems.	Transparency, black box	Compliance
6	The member company ensures that the behavior of the application can be actively monitored and that AI application users always have a means for providing feedback.	Transparency, feedback	Compliance
7	The member company contributes to sharing knowledge and providing education about AI in general, within and outside the sector.	Education	Compliance
8	The member company provides available information within the chain regarding: <ul style="list-style-type: none"> • the systems and the underlying use of the data • the technology applied • the learning curve for the system • data usage outside the EU 	Transparency, privacy	Compliance

A clear theme that can be observed in the code is the concept of transparency, which comes back in six of the eight principles. Most of these ‘transparency’ principles relate to providing certain information to users or other actors in the chain. This information can range from the responsibilities that different actors have, data that is used, information with regards to the

decision-making algorithm, the 'behavior' of the application, information regarding the technology, learning curve, and data usage outside the EU. These transparency principles are mostly about the responsibilities of actors in the use and application of AI-systems. Another strong focus of the current code is on privacy. Although the principles that contain privacy related elements (1, 4, and 8) are all relevant ethical principles for AI use and application, they reverberate the General Data Protection Regulation. These norms are relevant for IT-systems that use personal data in the 21st century but offer little novelty for AI design, use, and implementation.

Novel challenges for artificial intelligence arise due to automated decision making, the processing of ever-larger quantities of data, and the complexity of machine learning algorithms. These cause challenges such as fairness, explainability, accountability, and safety to play a new role outside of the realm of traditional information systems. However, it is precisely these challenges that the code of ethics fails to adequately address or even omits. First, even though explainability is briefly mentioned in a sub-sentence in the first principle, it does so in a very fuzzy way. Section 3.2.2.7 motivated the importance of explainability for AI-systems, hence one would expect this to be an important part of a code of ethics. If one finds the word 'ease of explanation' in the first principle, it is still unclear what is expected from an organization. With regards to explainability, the principle states: "the member company is aware of and anticipates the influence that implementation of AI can have on public values such as the principle of ease of explanation". This is a value-based statement that asks companies to be aware of explainability but offers little practical assistance for daily ethical decision making.

Fairness is another principle that the code of ethics currently does not address in a meaningful way for organizations in the industry. For fairness, the principle reads: "the member company is aware of and transparent regarding the state of the technique for AI and its technical possibilities but is also aware of the technical limitations of applications and the necessity to communicate clearly regarding this. This leads (among other things) to minimizing undesired "bias" and to promoting inclusive representation" (NLdigital, 2019, p. 1). This principle suggests that a causal relationship exists between being aware of the technical properties of AI and fairness. Namely, this awareness is supposed to minimize undesired bias and promote inclusive representation. The validity of this statement can be discussed, but such a statement does not provide 'guidance' to organizations within the industry. The principle urges companies to be aware of and transparent regarding the state of the technique for AI but provides no norm for what fair is, what is expected of organizations with regards to fairness, and how fair AI-systems can be achieved. Finally, section 3.2.2.6 motivated the importance of thinking about safety and creating non-maleficent AI-systems. Although we recognize that principles in the code could help towards the goal of a safe AI-system, there is no principle in the code of ethics that explicitly mentions this.

Concluding, the code of ethics is strong with regards to privacy and transparency of data and the system, but it fails to address explainability, fairness, and safety in an adequate way for the reasons mentioned above. Notions on both explainability and fairness should be extended in the code and should have an individual principle solely focused on demanding explainable and fair AI and dictating actions that can result in such systems. Furthermore, safety and non-maleficence should be added in a new principle so that organizations know and admit that they are expected to build robust and safe systems that minimize harm. These principles should be widely

applicable, but also offer practical guidance for decision-makers, in a similar fashion as the principles that demand the provision of a means for feedback for users and educational contributions to society by sharing the acquired knowledge about AI.

3.3.3 European Union ethical guideline

In 2018, the European Commission commissioned a high-level expert group on artificial intelligence to create a draft for a European ethical guideline for trustworthy AI (HLEG AI, 2019). Although the title of the guideline states trustworthy AI, experts treat it as a guideline for ethical AI. The high-level expert group on AI (HLEG AI) articulated four ethical principles that they treat as ethical imperatives. According to this deontological approach of the group, practitioners of AI should always seek to comply with these principles. The four principles are (HLEG AI, 2019):

- Respect for human autonomy
- Prevention of harm
- Fairness
- Explicability

Based on these four fundamental principles that the group deems important, seven key requirements for trustworthy AI were identified. An extensive ethical assessment follows these requirements in the document that was published by the European Commission. According to the high-level expert group on AI the following seven ethical requirements should be met in order for AI to be trustworthy (HLEG AI, 2019):

Table 10: Ethical requirements for trustworthy AI, adapted from: (HLEG AI, 2019)

Ethical requirement	Sub requirements
Human agency and oversight	Fundamental rights, human agency, and human oversight
Technical robustness and safety	Resilience to attack and security, fall back plan and general safety, accuracy, reliability, and reproducibility
Privacy and data governance	Respect for privacy, quality, and integrity of data, and access to data
Transparency	Traceability, explainability, and communication
Diversity, non-discrimination and fairness	the avoidance of unfair bias, accessibility, and universal design, and stakeholder participation
Societal and environmental wellbeing	Sustainability and environmental friendliness, social impact, society and democracy
Accountability	Auditability, minimization, and reporting of negative impact, trade-offs, and redress

By looking at this guideline and comparing it with NLdigital's code of ethics, we can identify room for improvement in NLdigital's code, and draw new insights for the adjustment and addition of principles. The ethical requirements that the European guideline demands are quite elaborate and encompass a wide range of sub-requirements. They also include the key principles that were identified in section 3.2.2.2, i.e. transparency, fairness, non-maleficence, responsibility, and privacy. Notable is that where NLdigital omitted safety and addressed explainability and fairness minimally, the EU guideline treats those three as ethical imperatives. They translated these principles into ethical requirements for the use of AI within organizations. These requirements

also have a heavy emphasis on fairness, safety, and explainability. One ethical requirement, for example, is technical robustness and safety. Under this requirement, the guideline demands resilience to attack and security, a fallback plan, and general safety, accuracy, reliability, and reproducibility. These requirements could be used as an example of the formulation of a safety principle for NLdigital’s code of ethics

3.3.4 Various enterprise ethics principles

In the previous sections, NLdigital’s ethics principles (3.3.2) and the European ethics guideline (3.3.3) were discussed in more detail. However, not only overarching organizations like the European Union and NLdigital but also many companies (e.g., Microsoft, IBM, Google, etc.) that are working with AI have published their own ethical principles. Especially, large multinational technology companies in the IT sector are increasingly showing these principles to the public. Although this research specifically focuses on including the experiences and struggles of small and medium-sized enterprises, looking at enterprise codes from large organizations is still important. These organizations have often put large amounts of research into the development of their ethical principles. Looking at them allows us to have a complete picture of the state of the current code of ethics landscape in the ICT-industry and to be aware of any industry-specific principles and requirements. Table 11 displays the ethics principles from IBM (2018), Microsoft (2019), and Google (2019), along with sub-requirements that this research found in explanations of these principles.

Table 11: Enterprise ethics principles of IBM, Microsoft, and Google

IBM	Microsoft	Google
Accountability <ul style="list-style-type: none"> Active responsibility 	Accountability <ul style="list-style-type: none"> Algorithmic accountability 	Accountability
Value alignment <ul style="list-style-type: none"> Alignment with values of users 	Inclusiveness	Socially beneficial
Explainability <ul style="list-style-type: none"> Transparency 	Transparency <ul style="list-style-type: none"> Understandable 	Uphold scientific standard <ul style="list-style-type: none"> Education
Fairness <ul style="list-style-type: none"> Biases Inclusion 	Fairness <ul style="list-style-type: none"> Biases 	Fairness
User data rights <ul style="list-style-type: none"> Protection Access 	Privacy & Security	Use privacy design principles <ul style="list-style-type: none"> Privacy Transparency
	Reliability & Safety	Safety Availability for applications that follow these principles

The principles are broadly consistent over the three companies. Nonetheless, one thing that stands out is the last principle that Google states, i.e. “AI be made available for uses that accord with these principles” (Google, 2019, p. 1). This means they are not only taking responsibility for their own AI practices but also take active responsibility for the practices of other actors that they

supply their technologies to. Google also formulated four application purposes for which they refuse to supply AI technologies, e.g. technologies and weapons that cause harm, and surveillance and information technologies that violate widely accepted norms. With regards to the addressing of the three challenges (i.e., safety, fairness, and explainability) that were previously identified as weak or missing in NLdigital's code, these companies have included them in their own principles. This means that they value these requirements and want everyone in their organization to take them into account when designing, implementing, or using AI-systems. The fact that these principles are addressed in ethical codes of large companies within the industry only further supports the need to assess whether these are impactful challenges for the entire industry (i.e., also small and medium-sized enterprises). In contrast to these big companies, SMEs do not have the time, money, resources, and infrastructure for dealing with their own code of ethics. At the same time, smaller companies in the ICT-industry desperately need NLdigital and their support. This underlines the importance of having a good code of ethics for small and medium-sized enterprises.

Sub-research question 2: What ethical aspects in NLdigital's code of ethics are missing that can be identified by comparing it to other codes of ethics?

Answer: Automated decision making, the processing of ever-larger quantities of data, and the complexity of machine learning algorithms causes challenges such as fairness, explainability, accountability, and safety to play a new role outside of the realm of traditional information systems. However, it is precisely these challenges that NLdigital's code of ethics fails to adequately address or even omits. Through analyzing NLdigital's code of ethics comprehensively we identify three 'weak' spots or areas for improvement. First, the principle of **explainability** is rather vague and does not cover off explainability sufficiently. It is a value-based statement that asks companies to be aware of explainability but offers little practical assistance for daily ethical decision making. Second, the principle of **fairness** currently also provides minimal guidance for businesses and is formulated ambiguously. The current principle suggests that a causal relationship exists between being aware of the technical properties of AI and fairness. The validity of this statement can be discussed, but definitely it does not provide a norm for what fair is, what is expected of organizations with regards to fairness, and how fair AI-systems can be achieved. Finally, section 3.2.2.6 motivated the importance of thinking about safety and creating non-maleficent AI-systems. Although we recognize that principles in the code could help towards the goal of a safe AI-system, there is no principle in the code of ethics that explicitly mentions **safety or non-maleficence**.

When comparing NLdigital's code of ethics with other codes of ethics and guidelines for AI, we find that explainability, fairness, safety, and non-maleficence are typically well-addressed in these. In the analysis of 84 guidelines and codes by Jobin et al. (2019), the most frequent appearing principles were transparency, fairness, and non-maleficence. Another important guideline for businesses in Europe that are working with AI is the European guideline for trustworthy AI. This guideline too, put specific emphasis on principles for fairness, transparency, and safety, with specific sub-requirements for explainability and interpretability. Lastly, we analyzed three different list of ethics principles from large multinational businesses (i.e., IBM, Microsoft, and Google). These too included at least two out of the three key principles (i.e., fairness, explainability, and safety).

3.4 Conclusion

The first part of the literature review (3.1) is dedicated to a review of the technical domain. This research focusses on the ethics of weak or narrow AI. Therefore, it does not consider topics such as superintelligence and singularities that are associated with strong AI and artificial general intelligence. The working definition that will be used throughout the rest of this research considers AI as machines, mostly computers, that are programmed to interpret external data, learn from such data, and use the learnings to successfully carry out a single task. These 'narrow' systems carry out this task within a pre-established range, even if they appear to be operating in a much more sophisticated way than that. Almost any form of AI that is used in practice by companies in the industry is powered by machine learning algorithms.

The second part of the literature review (section 3.2) was dedicated to setting the theoretical domain of this research by a review of the main views in normative ethics, the prominent works in ethics and artificial intelligence, and the factors affecting the success of codes of ethics. For the descriptive part of this research in the ICT-industry, a useful perspective is that of moral pluralism. This pluralistic approach acknowledges that different viewpoints exist in the industry. The interviews and questionnaire will show this plurality. However, in the pursuit of a better code of ethics, this research will defend specific priori agreeable moral values. Transparency, fairness, explainability, safety and non-maleficence, privacy, responsibility, and accountability were identified as important moral values through a thorough review of AI ethics literature (section 3.2.2). The importance of these values in the design and implementation of artificial intelligence has been shown and forms the starting point for identifying opportunities for improvement in the code of ethics.

Other important concepts resulting from the academic literature that will be used throughout the next phases of this research are the necessity of whistleblowing structures, pillars of ethical governance, and the influence of code quality on code success. First, different studies (Boddington, 2017; Bowden, 2008; Winfield & Jirotko, 2018) have highlighted the importance of anonymous channels for reporting misconduct in parallel with a code of ethics. Currently, NLdigital does not offer such a mechanism, nor does their code of ethics prescribe the necessity of such a mechanism. Empirical evidence from the interviews and questionnaire will need to show if companies in the industry have created such channels themselves, or that there is a gap between literature and practice present. Second, the framework of the five pillars for solid ethical governance by Winfield and Jirotko (2018) has been identified as useful in assessing the quality of ethical governance in organizations. This framework will be used in both the interviews and survey to assess where companies are in terms of ethical governance. This study expects large multinational corporations to possess more pillars than small and medium-sized companies due to a lack of resources. The results of the interviews and survey will inform the need for additional tools and mechanisms to ensure good ethical governance of AI within these different sizes of ICT-organizations. Third, code quality was established to have a significant impact on the success and adoption of a code (Erwin, 2011). Especially, the two heaviest weighing components in determining code quality, (i.e., are readability and tone, and the addressing of all relevant risk topics) will be key topics to assess and ask feedback on during the interviews and questionnaire.

In the last part (section 3.3), different ethics guidelines and codes of ethics were reviewed to form a source of inspiration and have a complete picture of the state of these documents in the ICT-industry. Furthermore, NLdigital's code of ethics was analyzed extensively. It is concluded that the code is strong with regards to transparency and privacy but lacks depth and practicality with regards to explainability and fairness. Moreover, it lacks the addressing of safety and non-maleficence altogether. The importance of safety and non-maleficence for AI was shown in section 3.2.2.6. An important goal of the descriptive data gathering will be to establish if these are important challenges for companies in the industry on a practical level and if these companies have tools and mechanisms to deal with them.

Finally, principles and values from different codes of ethics, guidelines, and ethical principles from a diverse set of organizations can be broken down into building blocks to form a conceptual overview. Table 12 displays the combined overview of the principles and values that were identified throughout the review of the different guidelines in section 3.3. This conceptual overview will form a framework for the semi-structured interviews. Ethical challenges that companies mention during these interviews should fall somewhere within this framework. The ethical building blocks and concepts in Table 12 are not hierarchical yet, thus no value is currently giving to the position in the overview. This is because literature typically does not provide clear rankings of importance, more than frequencies in meta-analyses (Jobin et al., 2019). The interviews and questionnaire, however, do aim to show a distinction in the importance of the different ethical challenges with AI for organizations in the ICT-industry.

Table 12: Conceptual overview of ethical challenges with artificial intelligence

Ethical building block	Important concepts	Source
Accountability	<ul style="list-style-type: none"> - Algorithmic - Auditability - Responsibility gap - Liability 	(Baylé, 2019; Bostrom & Yudkowsky, 2014; Google, 2019; HLEG AI, 2019; IBM, 2018; Kroll et al., 2016; Matthias, 2004; Microsoft, 2019; Mittelstadt et al., 2016; Russell & Norvig, 2016; Salvino, 2018; Smith & Browne, 2019)
Transparency	<ul style="list-style-type: none"> - Algorithmic - Trust - Understandability - Traceability - Explainability - Communication 	(Baylé, 2019; Bostrom & Yudkowsky, 2014; Google, 2018; HLEG AI, 2019; IBM, 2018; Microsoft, 2019; NLdigital, 2019; Smith & Browne, 2019)
Responsibility	<ul style="list-style-type: none"> - Awareness - Active - Responsibility gap 	(IBM, 2018; NLdigital, 2019; Yao et al., 2018)
Fairness	<ul style="list-style-type: none"> - Biases - Inclusion - Distribution - Discrimination 	(Baylé, 2019; Google, 2019; HLEG AI, 2019; IBM, 2018; Microsoft, 2019; NLdigital, 2019; Smith & Browne, 2019; Yao et al., 2018)
Privacy	<ul style="list-style-type: none"> - Protection & security - Access & control 	(Google, 2018; HLEG AI, 2019; IBM, 2018; Lee & Park, 2018; Microsoft, 2019; Pardo & Siemens, 2014; Smith & Browne, 2019)

Safety	<ul style="list-style-type: none"> - Security - Reliability - Continuous evaluation - Fall back plan 	(Google, 2018; HLEG AI, 2019; Lee & Park, 2018; Microsoft, 2019; Smith & Browne, 2019)
Education	<ul style="list-style-type: none"> - Knowledge sharing - Learn ethical theories - Equality 	(Goldsmith & Burton, 2017; Google, 2018; NLdigital, 2019; Yao et al., 2018)
Beneficial to society	<ul style="list-style-type: none"> - Sustainability - Social impact - Democracy 	(Google, 2018; HLEG AI, 2019)
Human agency and oversight	<ul style="list-style-type: none"> - Human autonomy - Oversight 	(HLEG AI, 2019)
Auditability	<ul style="list-style-type: none"> - Black-box testing - White-box testing 	(Bostrom & Yudkowsky, 2014; Kroll et al., 2016; Salvino, 2018)
Value alignment	<ul style="list-style-type: none"> - Desired values 	(IBM, 2018)
Predictability	<ul style="list-style-type: none"> - Governance/law 	(Bostrom & Yudkowsky, 2014)
Job disruptions	<ul style="list-style-type: none"> - Human-machine collaboration - Unemployment 	(Baylé, 2019; Russell & Norvig, 2016; Yao et al., 2018)
Incorruptibility	<ul style="list-style-type: none"> - Malicious AI - Robust 	(Bostrom & Yudkowsky, 2014; Russell & Norvig, 2016; Yao et al., 2018)
Feedback	<ul style="list-style-type: none"> - Whistleblowing - User feedback 	(Boddington, 2017)

4. Interview results

This chapter presents the findings from the semi-structured exploratory interviews that were conducted in accordance with the interview protocol (REF Appendix A). The analysis of the six interview transcripts was done in Atlas.ti. The coding of the document relied mostly on inductive coding as the questions were aimed at exploring current ethical practices and challenges with artificial intelligence in the industry. In total, 187 codes were identified based on relevance to the research question, some of which were overlapping. These codes were clustered in 23 categories that reflect the important topics of this research. The different categories and their subsequent codes are listed in appendix B. These categories will be discussed in the next sections where the most important and recurring codes are presented under different themes that reflect the research questions of this study; the landscape of AI, ethical challenges with AI, mechanisms for increasing ethical behavior and feedback on the code of ethics. The structure of the chapter follows the interview protocol. Quotes and information by the different interviewees (i.e., IV1, IV2, IV3, IV4, IV5, and IV6) are presented to underline the findings of the interviews.

With regard to AI as a technical artifact, the aim of the interviews is to understand how and for what AI is currently used in the sector (section 4.1). With regard to ethics, the interviews aim to explore which ethical challenges are encountered and seen as important in the industry (section 4.2). Moreover, the interviews explored which mechanisms organizations currently leverage to increase ethical behavior with artificial intelligence (section 4.3). Lastly, feedback on the current code of ethics was collected to understand its strong points and shortcomings and identify opportunities to enhance its impact (section 4.4). Finally, key ethical building blocks that together with the literature review form the foundation for the construction of the survey, are identified (section 4.5).

4.1 The landscape of Artificial Intelligence

All interviewed companies use forms of machine learning. IV4 immediately requested at the beginning of the interview to talk about machine learning for the rest of the interview instead of artificial intelligence. This underlines that the predominant form of artificial intelligence used in ICT-businesses today is machine learning. IV5 described that the industry is structured as a service, and from the interviews, it is indeed understood that almost all interviewed companies provide some sort of service. Half of the interviewed companies that use machine learning provide a very concrete service to customers in more of a consulting sense. IV1, IV3, and IV5 indicate that customers often have a problem with their data. Many companies possess large quantities of ‘unused’ data but don’t have the talent or other resources to leverage this data. These kinds of companies will often approach ‘AI-consultants’ such as IV1 and IV3 for a short time to optimize their product or service using the data that they have. These ‘consultancy’ service providers range in size from small national firms to large multinationals. IV5 solves the same problem of unused data for companies by providing them with a low entry machine learning platform. Their platform uses open-source libraries to give analysts without a degree in machine learning or statistics the capabilities to use and develop their own AI models. IV2, IV4, and IV6 provide an AI service in a slightly different sense. They use machine learning to automate the analysis of a text. They then filter the text and provide relevant information to their customers or users. IV2 uses this to flag

risk areas in acquisition documents, IV4 to rank information and filter harmful or unwanted content, and IV6 to recommend relevant political information.

The majority of the technical fundamentals of the Machine learning ecosystem are open source, e.g., libraries such as Scykit and Keras. This makes it possible for companies without extensive AI knowledge but with knowledge of certain data or software development to use machine learning to their advantage (IV2, IV3, and IV6). Specifically, it is clear that large multinationals companies develop almost all their technologies in-house. They employ some of the best AI talents and have the resources to research and develop their own infrastructure and systems. These organizations also already have established AI practices. On the other hand, SMEs, as IV2, IV3 and IV6 indicate that even though they already actively use machine learning they are still in the development phase. Some of them develop their own software using free libraries such as Scykit (IV5), whereas others use tools, plugins, and platforms such as that of IBM's Watson (IV6) or Microsoft Azure (IV2, IV3).

4.2 Important ethical challenges with AI in the industry

The interviews focused on a large part of the ethical challenges that the interviewed companies experienced or frequently have to deal with in their daily work with AI. The quality of the technical artifact is for SMEs usually the priority (IV2). Ethical challenges arise mostly after a system has been developed, or at least this is when companies actively start to think about them. They start thinking about how they can design mechanisms to prevent such challenges into the system itself or develop supporting processes to prevent or approach the issues. Ethical challenges are, therefore both important in the design of systems as well as in the implementation of products or services.

The first part of this section (4.2.1) displays the ethical challenges that companies judged as most pressing to them and their business. Section 4.2.2 is dedicated to the various other practical ethical challenges that the companies experience, and section 4.2.3 discusses the level of homogeneity of these challenges among colleagues within the same company.

4.2.1 Most pressing ethical challenges

During the interviews, the respondents were asked to list the most pressing ethical challenges that they encounter in the use of artificial intelligence. One interviewee (IV6) mentioned only one key challenge, whereas all the others mentioned at least two impactful challenges. Table 13 displays these challenges thematically ordered by occurrence as analyzed through atlas.ti along with important keywords that were mentioned in the context.

Table 13: Most pressing ethical challenges resulting from the interviews

Challenge theme	Keywords	Occurrence	Interviewee
Fairness	Bias, equality, evidence, context dependence	4	IV1, IV3, IV4, IV5
Explainability	Transparency, level of understanding, complexity, customer demands, context-dependence	4	IV1, IV2, IV5, IV6
Safety	Unexpected outcomes, wrong predictions, non-algorithmic, open-source, harm	2	IV3, IV4
Privacy	Data collection, GDPR, privacy up to clients, data governance, customers demand	2	IV1, IV2

4.2.1.1 Fairness

Fairness was named by four interviewees as one of the most pressing challenges. Challenges with fairness mainly relate to the mitigation and prevention of unwanted biases. These biases may result in discrimination based on, for example, gender and ethnicity, as well as an unfair allocation of services and their corresponding prices. Models can reflect biases of the people that are working on the system, but a novel difficulty with AI is that when biases are present in the training data, they will also creep into the ML-model. This is something that the companies that are working with personal, or human data actively have to deal with. The difficulty according to IV5 is that fairness in terms of biases is strongly dependent on the context. IV5 said: “How do you tell people, if all your features are related to race or ethnicity, that’s a bad thing? Sometimes it might also not be. If you are thinking about a disease prediction model, then these labels are fair game. Those things actually represent key things about a person’s identity or background. But if you’re making a loan prediction or loan default prediction model, then you should probably not be using those”.

4.2.1.2 Explainability

The second key challenge in the industry, as indicated by the interviews, is that of explainability and interpretability. An interesting finding is that all participants immediately named explainability rather than the wider concept of transparency. Explainability has become a pressing issue recently because the industry and customers demand this from service providers. The key difficulty with the explainability of AI is that it is very complex to explain the system or decision. First of all, this is hard because the level of understanding of clients, customers and consumers is much lower than that of designers and providers of AI-systems. These designers, in this case, companies in the ICT-industry, are finding it very difficult to find the right level of abstraction in explainability. Second, ICT-companies have to carefully balance the things they want to explain to their customers with the time they can ask them to understand it. If they ask too much time, the client will become dissatisfied and companies will eventually lose them. A third reason why it is complex to explain AI-systems is that they are, to some extent, a black-box. AI-systems may teach themselves to recognize new patterns, and deep learning models become so complex to

the extent that even the creators do not fully understand them anymore. Clearly, explaining the process to someone with limited knowledge will then be a complex undertaking. Lastly, explaining systems or decisions is difficult because it is largely dependent on the context. IV4 indicated that you cannot always be transparent and explain decisions because sometimes information is confidential or people could reverse engineer explanations back to individuals.

4.2.1.3 Safety

Safety is a difficult challenge for companies because of its unexpectedness and unpredictability. IV3 uses an example of a project in which they are predicting waiting for patients in hospitals to demonstrate how wrong predictions may cause harm. Although they try to build as accurate prediction models as possible, there is always the possibility that predictions are wrong and endanger the safety of people. For IV4, safety and non-maleficence are really about preventing society from harm. They do not only consider the consequences of fake predictions but also try to actively prevent people from harm as a cause of the existence of their platform, even if this falls outside of the scope of the results of their own actions. For example, IV4 says that they take an active role in preventing suicide by putting psychological experts on suicide-related messages that are posted on their platform. Another characteristic that makes safety and non-maleficence with AI an issue is an open-source ecosystem that it exists in (IV5). Almost anyone can now make AI-system or use a large number of different vendors to create a system, and eventually, no one will know what is going on. This makes it hard for AI/ML service-providers to ensure if their products are used in a non-maleficent way.

4.2.1.4 Privacy

Lastly, privacy is an important challenge for companies in the industry. For IV1, the most ethical challenges for privacy arise in the data collection phase, for example, when you are collecting data that you are not supposed to, when you are dealing with sensitive data, or when you are combining different databases and that combination could be dangerous because it could lead to a specific individual. Confidential or secret data of clients is also an issue for IV1. An important challenge for them is to develop and train their system with this kind of data that they're not allowed to look into. Another challenge is that service providers can often not impose extra data governance rules upon clients, as this will cause them to lose their clients. Most companies, therefore, limit their privacy practices to comply with the GDPR. According to IV1 the GDPR is a sword with two sides, on the one side, it is great because it prevents abuse but it also makes the work of data scientists harder because some information that they would benefit from having, falls under the GDPR, and you are not allowed to collect it.

Sub-research question 4: Which ethical challenges are most pressing to Dutch ICT-businesses in the development and application of AI-systems?

Partial answer (part 1): Fairness and explainability were both named by four interviewees as the most pressing challenges that they encounter in the development and application of AI-systems artificial intelligence. Challenges with fairness mainly related to the mitigation and prevention of unwanted biases. These biases may result in discrimination based on for example gender and ethnicity, as well as an unfair allocation of services and their corresponding prices. Models can reflect biases of the people that are working on the system, but a novel difficulty with AI is that when biases are present in the training data, they will also creep into the ML-model. The second key challenge in the industry as indicated by the interviews is that of explainability and interpretability. Explainability has become a pressing issue recently because the industry and customers demand this from service providers. The key difficulty with the explainability of AI is that it is very complex to explain the system or decision. First of all, this is hard because the level of understanding of clients, customers and consumers is much lower than that of designers providers of AI-systems. These designers, in this case, companies in the ICT-industry, are finding it very difficult to find the right level of abstraction in explainability. Second, ICT-companies have to carefully balance the things they want to explain to their customers with the time they can ask them to understand it. A third reason why it is complex to explain AI-systems is that they are in fact to some extent a black-box.

Challenges that were named less frequently by the interviewees but we still consider to be significant are safety and privacy. Safety is a difficult challenge for companies because of the unexpectedness and unpredictability of negative occurrences. Furthermore, once a technology is transferred to clients, it is hard for AI/ML service-providers to ensure if their products are used in a non-maleficent. Lastly, privacy is an important challenge for companies in the industry. The most ethical challenges for privacy arise in the data collection phase, for example, when businesses are collecting data that they are not supposed to or when businesses are dealing with sensitive data. In the end, most companies indicated to limit their privacy practices to comply with the GDPR.

4.2.2 Other ethical challenges

Apart from the four challenges that were identified in the previous section as the most pressing ethical challenges, interviewees have described other ethical challenges in their work with AI. First, responsibility is still a relatively immature topic for many of the companies in the industry. Most of the interviewees recognize that responsibility is an ethical challenge worthy of discussion in their practice, but not something they currently are actively thinking about. This proves the urgency of adopting responsibility or accountability in a code of ethics. IV1 and IV4 said that responsibility is mainly important for very sophisticated AI-systems or systems with a big societal impact. However, they state that in most cases responsibility is currently not really an issue because of the small decisions that are being made. IV2 and IV3 both describe the importance of discussing responsibility, but also acknowledge it is not a topic that they actively talk about, nor that they know how to deal with. IV5 talks specifically about the responsibility gap in practice, and the difficulty for them because they don't know what their client uses their tool for. SMEs are careful in addressing sensitive topics as these because it may cause them to lose clients (IV5).

This is a point that is important for other ethical challenges as well, SMEs are limited in the demands they can make to customers as these don't want them to interfere in their affairs.

The default for all the interviewed companies is that responsibility lies with the client or customer. Companies are delivering a service to others based on their requirements, data, and care of the solution after implementation and believe that the customer is therefore eventually responsible for the consequences it may cause. Similar argumentation was given by providers of AI-assisted tools (IV2, IV5, and IV6), who claimed that their product is just a tool to make people's practices easier. They believe customers should be aware that they are using a tool and that they are responsible for whatever they build it for. We believe that this approach of neglecting ethical responsibilities by companies is not representative of good ethical AI practices. It further underlines the need for this research and the need for a better code of ethics and additional concrete measures that make companies aware of the responsibilities that they own.

4.2.3 Interpretation of colleagues

In the interviews, interviewees were asked if they believe that their colleagues saw these ethical challenges and the way in which their organization should deal with them differently. The larger organizations mentioned that ethics are really embedded in their organization and that they know what is expected of them. These types of organizations have a set of ethics principles (IV1), on top of that they cannot afford to neglect their ethical responsibilities because they are under strict public scrutiny (IV4). Moreover, these companies have an open culture of discussing ethics and employ numerous experts for relevant fields of knowledge that they could ask for help. Employees in these organizations are generally seeking this help and are willing to adapt to different viewpoints.

In SMEs however, the interviews show that the interpretations of colleagues are less homogenous. IV3 does not really discuss ethics within their organization and describes that there may be misaligned ethical thinking within their organization. Others sometimes discuss ethics with their colleagues but in a more informal way. Not on any occasion, a set of formal ethics principles was present to streamline ethical norms and values within SMEs. IV2, IV3, IV5, and IV6 all four indicated that they have some sort of unwritten understandings with colleagues or a list of personal principles, but in no company are these values officially streamlined. IV5 even explicitly mentioned that currently, the individual employee values are leading. Companies all seem to trust that their colleagues want to create as 'ethical' systems and products as possible, but what that means for others is ambiguous and done in good faith.

Sub-research question 3: How do companies in the Dutch ICT-industry perceive the ethical challenges of AI?

Answer: Throughout the interviews we have tried to understand the differences in judgment and perception of the ethical challenges and the way in which the industry deals with them. First, companies mainly think of fairness as the mitigation and prevention of unwanted biases. These biases may result in discrimination based on for example gender and ethnicity, as well as an unfair allocation of services and their corresponding prices. For explainability and interpretability, the key difficulty for businesses is the complexity to explain the system or decision rather than being 'transparent'. This is hard because the level of understanding of clients, customers and consumers is much lower than that of designers and providers of AI-systems. Furthermore, companies perceive they have to carefully balance the things they want to explain to their customers with the time they can ask them to understand it. Finally, businesses cannot always be transparent and explain decisions because sometimes information is confidential or people could reverse engineer explanations back to individuals. In any case, businesses understand the importance of explaining the AI's decisions to customers and the public. For most businesses, safety is really about preventing society from harm. Some go to the extent that they do not only consider the consequences of 'bad' predictions or decisions but also try to actively prevent people from harm, even if this falls outside of the scope of the results of their own actions. Still, all businesses in the industry try to build as accurate prediction or decision models as possible. Furthermore, companies that provide AI or ML as a service find it difficult to ensure if their products are used in a non-maleficent way. Lastly, in terms of privacy, businesses mostly perceive behaving 'ethical' as complying their privacy practices with the GDPR. Businesses, especially service providers, believe they cannot impose extra data governance rules upon clients, as this will cause them to lose their clients.

However, these perceptions are not ubiquitous. Some companies for example indicate to discuss ethics with colleagues and are willing to adapt to different viewpoints (e.g. their colleagues). In SMEs, the interviews indicate that the interpretations of colleagues are not very homogenous. Ethics is less frequently discussed within their organization and there is a high probability of misaligned ethical thinking within their organization. Not on any occasion, a set of formal ethics principles was present to streamline ethical norms and values within SMEs. Currently, the individual employee values are leading. Companies all seem to trust that their colleagues want to create as 'ethical' systems and products as possible, but what that means for others is ambiguous and done in good faith.

4.3 Mechanisms for dealing with ethical challenges and increasing ethical behavior

This section discusses the mechanisms and activities that organizations have in place to encourage, enforce and assist ethical behavior with artificial intelligence for their employees. Sub-sections 4.3.1 and 4.3.2 describe the mechanisms for dealing with specific ethical challenges and encouraging ethical behavior with AI in general respectively. Subsequently, sub-section 4.3.3 discusses the pillars for good ethical governance of AI (Winfield & Jirotko, 2018) that are present in the companies of the interviewees.

4.3.1 Mechanisms for dealing with ethical challenges

4.3.1.1 *Fairness*

Companies that indicated they are experiencing challenges with fairness in their work with AI have relatively well-developed tools. IV1, IV4, and IV5 all indicated that they use software kits, which are mostly self-developed, to test for fairness in their data. Specifically, these tools are aimed at testing biases in sub-populations so that these can be minimized. IV3 also actively tested their data and results on intrinsic bias towards sub-populations but did so manually. All the companies stated that this is a good method to prevent unwanted bias that may result in an unfair system. Another mechanism to increase fairness was used by IV4. He/she indicated that their company is always aiming to have diverse teams consisting of people from different age groups, backgrounds, and looking from different perspectives. IV4 believed that together with a good understanding of your data, a diverse team can bring you fair machine learning and artificial intelligence.

It becomes clear that most interviewees were unable to point out what particular notion of fairness is suitable for them. However, their organizations still adopted different tools and mechanisms to minimize or prevent biases. It seems they adopted standard tools that they have been told they are fair, rather than have a thorough understanding of what a fair AI system entails.

4.3.1.2 *Explainability*

As established in section 4.2.1, transparency is predominantly interpreted as providing the possibility for customers to understand the results and decisions made by AI-systems. The interviewees mentioned different ways in which they try to explain their system or make the systems more explainable. Some interviewed organizations (IV4, IV5, and IV6) are trying to build the ability to explain to users what is happening in their software tools. This can be in the form of a configurable filter (IV6), a list of 'reasons' for why you get a certain recommendation (IV4), or a complete log and documentation of everything that has happened in the system (IV5).

Another way in which companies in the industry try to make their systems explainable is by actively providing information to their users (IV1, IV3, and IV5). IV1 and IV3 focus their efforts around educating clients and explaining their model in a continuous client – company interaction. Furthermore, on top of education, IV5 also creates tutorials for clients that they can follow to really understand all the systems that they're using.

4.3.1.3 *Safety*

With regard to safety, companies always try to build the most accurate and robust models and systems. However, that does not rule out any harm or other negative consequences. According to IV4, the danger with safety lies in the unforeseen impact that an AI-system might have. Companies in the industry try to minimize the occurrence of such unforeseen consequences in different ways, but don't build scenarios of what to do in case of these consequences. IV3 indicated that building such a safety net is something they should do in the future, but currently, they are not actively doing this. The company uses an example of a project in which they estimate waiting times in hospitals to admit that it's a bit of a grey area about what happens in case of a misprediction.

As mentioned before, some companies are aware of the things they can do to prevent or minimize the occurrence of harmful consequences. IV4, for example, tries to minimize these consequences by always having a human check build into the loop as well. His/her organization also uses machine learning algorithms as a tool to increase safety and non-maleficence by filtering hate speech and suicide-related messages, and then, for example, putting psychological experts on it in the pursuit of suicide prevention. Like IV4, IV6 also actively validates the results of their systems using frequent random checks on the results that customers get to see. On top of that, they are very selective in the use of their data sources, e.g. they use primary data sources from government organizations. IV5 approaches the safety challenge differently because they don't actually develop end-applications and have no insight into the data that customers plug into the platform. They, therefore, focus on checking the legitimacy of the client. At different points during the engagement with the client, they will validate if the client is a 'safe' or an ethical actor.

4.3.1.4 Privacy

All the companies that were interviewed operate in the Dutch ICT-sector and all collect or use data of European citizens. This means they all must comply with the General Data Protection Regulation (GDPR). Compliance with the GDPR is one way in which all interviewees dealt with privacy challenges. Some companies limited their privacy-related activities to those that are described in the GDPR. IV3, for example, indicated that they anonymize data, have data processing agreements, delete unnecessary data, but don't do more on privacy than is required. IV1 as well indicated that they don't know what to do more about privacy than the existing regulations require. The three largest organizations take privacy-enhancing activities further than just compliance with the GDPR. IV1 does two things; 1) they always make sure that they know how the data of a client was acquired. If they believe the data was acquired illegally in any way, they will decline the project. 2) the organization organizes training for everyone in the organization on the topic of privacy to carefully explain to them what kind of information they can and cannot look at. IV 4 describes that they have installed numerous privacy checks throughout the development and lifetime of products. In these checks, they will assess how they are doing from a privacy standpoint, and how the data privacy of customers can be better protected. Lastly, IV5 indicated the use of a slightly different mechanism to deal with privacy challenges. As mentioned earlier, IV5 doesn't see or possess any of the data that clients use to develop ML tools on their platform. Instead, they go beyond the scope of their own required data privacy activities and actively help their customers. IV5 describes that their organization helps its customers with their data operations, e.g., they help the customer organization to create data stewards. These are central figures in a companies' data management process that concern themselves with the governance of and collaboration around data.

4.3.1.5 Responsibility

As described in section 4.2.2, both SMEs, as well as large incumbents, put the end responsibility and accountability for unforeseen harmful consequences with their customers. Only two interviewees indicated ways in which they try to stimulate responsible AI. First, IV4 created clear community standards together with its customers. The standards describe what is accepted and not accepted information to share on their platform. This way they try to avoid the spear of malicious information. Second, IV5 tells that in their organization they are vocalizing responsible AI. IV says: "we try to focus on things like responsible AI, that is something that our CEO talks

about. We try to vocalize it and talk about it in terms of the concepts and discuss what that would mean.”

4.3.2 Internal mechanisms for increasing ethical behavior

In the interviews, interviewees were asked to describe the mechanisms that they have in place to increase or enforce ethical behavior. Some mechanisms were answered directly to this question, whereas some others were identified by the interviewer throughout the interview. The range of mechanisms and activities that were mentioned in the interviews are displayed in Table 14 below.

Table 14: Mechanisms and activities for increasing ethical behavior in the ICT-sector

Mechanisms and activities	Description	IV
Feedback mechanism for users	Customers can provide feedback on the selection criteria and the results that they are being shown. Often, they can immediately change certain variables.	IV4, IV6
AI ethics training	All employees working with AI get trained on the AI ethics principles of the company. Furthermore, the organization offers numerous different training modules on the topic.	IV1
Check the legitimacy of the client	During the engagement process with the client different teams (e.g., sales, and legal) check if the clients are responsible actors. Different background checks are done, some of these happen automatically.	IV5
Open culture about AI ethics	People can speak openly and freely about AI ethics and the worries that they have within the organization.	IV5, IV1
Whistleblowing channel for AI	Multiple channels where people within the organization can report misconduct. There is also the possibility to do this in an anonymous way.	IV1
Involve a wide range of stakeholders	Actively a wide range of stakeholders including customers, academia, government, NGOs and advocacy groups during the life cycle of a product.	IV1, IV4, IV6
Ethics principles and codes	Publish and adhere to your own or existing ethics principles or codes of ethics for artificial intelligence. These provide a guideline and checklist for engineers.	IV1, IV2
Publications and documentaries	Publish documentaries, articles, and information about AI and ethics to create public awareness and awareness by the clients.	IV5
Ethical leadership	Choose leaders partially based on their match with establishes ethics criteria. Advancement in the company will be determined on ethical behavior.	IV1
Media scrutiny	Some companies know that everything they do is under a magnifying glass by the media, and so they have to behave ethically to avoid public scrutiny.	IV4
Peer-reviewed work	You never or seldom go to a client alone, and you always have your work peer-reviewed. This is a mechanism to prevent bypassing of rules and violation of company norms.	IV1

Punishments or firing	Employees that behave in an unethical manner will be punished with lower bonuses, salary raises, or job opportunities. There is also a possibility that employees will be fired.	IV1, IV4
Collaboration with client	Discuss ethics and ethical challenges before and throughout the project with customers. Make sure the client is always on board by working in a client – company loop.	IV1, IV5, IV6
Communities and colleagues	Actively seek advice and support within (AI) ethics communities in the company and from colleagues that have dealt with the same or similar challenges before.	IV1
Central storage for ML models and data	All machine learning data sets, tools and models are stored in one central location. This allows for quick checking and review of the artifacts, as well as provides for the possibility to scale quickly	IV4
Feedback loop with ethics checks	The use of a frequent feedback loop in the creation of new products or services. A distinct part of the feedback should be attributed to an ethics review. It generally consists of internal stakeholders, but may sometimes also include external stakeholders.	IV4

4.3.3 Pillars of ethical governance

Interviewees were asked to indicate which pillars for good ethical governance are present in their organization. These pillars follow the framework that was proposed by Winfield and Jirotko (2018) and previously described in section 3.2.3.2. The presence of the pillars in the interviewed companies is shown below in Table 15.

Table 15: Presence of pillars for good ethical governance in the interviewed companies

		Interviewees					
		IV1	IV2	IV3	IV4	IV5	IV6
Pillars	1. Publish an ethical code of conduct	✓					
	- Alongside a whistleblower mechanism	✓					
	2. Provide ethics and RI training for everyone	✓					
	3. Practice responsible innovation	✓			✓		
	- Engage a wide range of stakeholders	✓			✓		✓
	- Undertake ethical risk assessments	✓			✓		
4. Transparent about ethical governance	✓						
5. Really value ethical governance	✓			✓			

There is a clear trend visible in the presence of pillars when contrasted with the size of the company. IV1 and IV4 are both interviewees from large multinational companies in the Dutch ICT-industry. These companies have vastly more ethical pillars on which they built their ethics practices than the SMEs in the industry. The majority of the SMEs indicate that they don't have any of the ethics pillars present in their organization. No SME that was interviewed publishes a code of ethics or a set of ethics principles, nor does any of them have a whistleblower mechanism

or channel to report misconduct in place. IV2 contributes the latter to the fact that the company does simply not employ enough people to add this function. Another possible explanation for the lack of such a mechanism is that the frequency in which such a channel will be used in SMEs is much lower than in large organizations. This research recognizes that it is understandable that SMEs have fewer pillars present because of the smaller resources and often the earlier stage that they're in, but that this is not something to strive for. Additional mechanisms, tools, and structures will need to be developed by and for SMEs to ensure solid ethical governance of AI in the sector.

Sub-research question 5: What concrete measures do Dutch ICT-businesses currently use to deal with the ethical aspects of developing and applying AI-systems?

Partial answer (part 1): One focus of the interviews was to map out the different measures that Dutch ICT-businesses use to deal with the ethical aspects of developing and applying AI-systems. We look at three different classes of measures and mechanisms, i.e., challenge specific measures, internal measures to encourage and facilitate ethical behavior, and pillars of ethical governance. First, we discuss different challenge specific measures that Dutch ICT-businesses use for their two most pressing challenges:

1. Fairness

- Software kits to test for fairness in data, specifically for testing biases in sub-populations.
- Manual checking of the data for intrinsic bias towards sub-populations.
- Working in diverse teams consisting of people from different backgrounds, ages, gender, ethnicity, etc.

2. Explainability

- Build the ability to explain to users what is happening into the software tools or services, e.g., a configurable filter, a list of 'reasons' for why you get a certain recommendation, or a complete log of the system.
- Actively providing information to their users
- Educating clients and explaining their model in a continuous client – company interaction.

Concrete internal measures to encourage and facilitate ethical behavior include: a feedback mechanism for users, AI ethics training, a check off the legitimacy of the client, an open culture about AI ethics, a whistleblowing channel for AI, ethics principles and codes, publications and documentaries about ethics, ethical leadership, media scrutiny, peer-reviewed work, punishments or firing, collaboration with the client, internal communities, a central storage for ML models and data, and a feedback loop with ethics checks.

Finally, we found the pillars of ethical governance by Winfield & Jirotko (2018) to be mostly absent in the interviewed SME sized businesses. Larger organizations on the other hand do commonly have ethical governance measures in place, e.g. publish a code of ethics, have a whistleblower channel, provide ethics and RI training, undertake ethical risk assessments, and engage a wide range of stakeholders.

4.4 Feedback on the current code of ethics

During the final part of each interview, participants were asked to look at the code of ethics for the Dutch ICT-industry that was published for NLdigital. They were then asked to provide their feedback on the code. An interesting first finding was that although most of the companies had heard of the existence of the code, only one company (IV2) was familiar with some of the content of the code of ethics. Still, all interviewees demonstrated a strong sense of reception for the code of ethics. Each interviewee provided their individual feedback on the code. This feedback is clustered below in two sections; section 4.4.1 presents the positive feedback on the principles that participants deemed 'good' and useful, and section 4.4.2 the feedback on the shortcomings of the current code of ethics.

4.4.1 Usefulness of principles

The feedback of the interviewees was univocal positive about the eight principles that currently make up the code of ethics for the industry. IV1 states that all the principles intuitively make sense and that anyone should strive for the principles that are listed. However, interviewees indicated multiple times that context-dependence makes some principles more applicable to companies than others. Interviewees mostly indicated principles that they were actively adhering to in their business. IV2, for example, valued the 5th principle most that describes the provision of the possibility to trace recommendations. IV3 and IV6 valued the 7th principle that prescribes the contribution of sharing knowledge about AI as they were both actively doing this. IV4 felt strongly about the feedback principle and indicated that this is something that they are incorporating into all of their products and services. Lastly, IV5 mentioned that their organization is actively working on the possibility to trace the recommendations or decisions made by systems on their platform, and therefore valued the 5th principle most as well.

4.4.2 Missing principles

Although the existing principles intuitively made sense and are all relevant to strive for according to participants, some of them were seen as fuzzy or limiting in guide-ability. Furthermore, participants indicated some important principles from their practice that they believe to be missing in the current code of ethics.

Most frequently, the interviewees mentioned that there is no 'good' principle on fairness included in the code of ethics. Although the word bias is mentioned once in the code, namely in the third principle of the code of ethics: "is aware and transparent regarding the state of the technique for AI and its technical possibilities but is also aware of the technical limitations of applications and the necessity to communicate clearly regarding this. This lead (among other things) to minimizing undesired bias and to promoting inclusive representation" (NLdigital, 2019, p. 1), the majority of the interviewees indicated that this was insufficient to provide any value for their practice. IV2 indicates that this principle only says to be transparent but fails to address any direction about what you can and cannot do. IV3, too, expresses criticism towards the address of fairness in the current code. IV3 says the code fails to address here the practical challenges in the work of a data scientist. Specifically, the interviewee says that the causal relationship in the principle is insufficient and offers AI practitioners no guidance as to what to do. IV5 also brought up the fact that this principle was very fuzzy. Apart from the principle offering little guidance, IV5 also says that no one really knows what fairness, equality, or inclusive representation means. The

interviewees suggest that a principle addressing the practical challenge of analyzing and testing data with the goal to minimize bias, and a notion of what fairness means should be included in the code of ethics.

With regard to fairness and inclusivity, IV4 indicates that another principle is missing that they take very seriously in their organization, namely the inclusion of a diverse group of people. Rather than addressing the minimization of bias, this is a principle around people and data. IV4 proposes the following principle: “Strives actively to work with a diverse group of people and stakeholders, including people from different backgrounds, ages, etc., with the goal to incorporate ethics and inclusivity in the design of the product or service.

Another principle that was fuzzy according to IV5 in the code of ethics is the first principle, specifically around the theme of ease of explanation. Explainability is one of the key challenges with AI, and this needs to be better addressed. IV5 and IV2 both miss information about the target group of who to explain your system or model to and the level of detail that should be included. Should the decision by the system be understood by data analysts/scientists or people with no knowledge of AI whatsoever? Finding this balance between interpretability and the level of detail is an important challenge that companies struggle with.

Moreover, some principle or guideline around a feedback loop is missing. One principle should refer to a feedback loop in ethics by the design process. IV4 says that when you start to design your system you have a certain application in mind, but this application may eventually change. You must have a continuous feedback loop to stay aware of new and evolving ethical challenges. IV4 uses an example of fake news that wasn’t an issue before but became a significant challenge for their organization in recent years. Through feedback loops and continuous ethical checks, they identified the evolving risks and created additional tools to deal with those.

Lastly, IV3 missed preventive privacy intrusion principles in NLdigital’s code. Many principles are about being transparent in the way you use data, but none of them tell you what good practice is around data privacy. IV3 thus suggests to add information such as use as many anonymous data as possible and don’t collect more data than you need.

4.4.3 Formulation of the principles

The opinions among the respondents about the formulation and the wording of the codes are mixed. Both IV1 and IV6 indicated that they like the formulation of the code and that the principles are easy to read and understand. IV3, IV4, and IV5 indicate that many of the principles are very abstract and are more about creating awareness around a certain topic than actually guiding in what one should do. IV3 expresses that this is not necessarily a bad thing, and is happy with the level of abstraction of the code. IV4, on the other hand, thinks that the principles are too abstract and that a more concrete layer should be added to the current principles.

Sub-research question 6: What are the common concerns that are uttered by the Dutch ICT-community about the current code of ethics offered by NLdigital?

Answer: Interviewees raised concerns about the usefulness, completeness, and formulation of the code of ethics. Although the existing principles intuitively made sense and are all relevant to strive for according to participants, some of them were seen as “fuzzy” or limiting in guide-ability. Most frequently, the interviewees mentioned that there is no ‘good’ principle on **fairness** included in the code of ethics. The interviewees indicated that the current principles are insufficient to provide any value for their practice in terms of fairness challenges. Interviews indicated that the current part on fairness does not provide any direction about what you can and cannot do. Furthermore, the interviewees state that the causal relationship in the principle is insufficient and fails to address here the practical challenges in the work of a data scientist. With regards to fairness and inclusivity, businesses are also concerned that a principle demanding the inclusion of a diverse group of people in the development and application of AI is missing. Rather than only addressing the minimization of bias, they thus also miss is a principle around people and data.

Another principle that was unclear according to businesses is that of the **ease of explanation**. Businesses are concerned about missing information about the target group of who to explain your system or model to and the level of detail that should be included. Finding this balance between interpretability and level of detail is an important challenge that companies struggle with. Lastly, an important principle that businesses believe to be missing in the current code of ethics is one around **safety**, having a safety net, and putting AI-systems to non-maleficent use.

4.4.4 Adoption and impact

In the final question of the interview, respondents were asked how the impact and adoption of the code of ethics can have more impact on their organization and for the industry. According to the larger organizations (IV1, and IV4), the code can mainly have a large impact on smaller organizations that don’t yet have ethical principles and are in the early stages of thinking about ethics. The interviewed SMES indeed show a willingness to adopt an industry-wide code of ethics. However, the code is currently insufficiently accessible (IV2), therefore, more communication and education should be done around the code (IV5). The code needs to be shared not only to an organization but also must diffuse across the entire AI practice of the organization. The code is now too often stuck with the leadership of ICT-companies. IV2 and IV6 also indicate that training around the code of ethics and the principles that it entails would be valuable for their organization. These smaller organizations currently don’t have the knowledge and resources to organize those training themselves. For this reason, reviewing the current code of ethics, determining the need for additional concrete mechanisms such as pieces of training, and performing these interviews also as a way to make companies know about them is of paramount importance.

Other possibilities that interviewees indicated to increase the impact of the code are:

- To make a certification from or around the code. The code should become a quality mark that can be granted or revoked if you do or do not handle in accordance with the code (IV2, and IV6).

- The code should be more concrete or have an extra concrete layer under it. This can, for example, be done by including an ethical checklist (IV5).

4.5 Conclusion

A total of six semi-structured interviews has been executed in order to gain an in-depth understanding of the AI and ethics practices of organizations within the ICT-industry. The different ethical challenges, tools, and mechanisms that organizations use have been explored, and feedback on the code of ethics has been gathered. The most notable findings from the interviews form a major factor in the development of the survey. Below we summarize the most important findings of the interviews that will be adopted for the design of the questionnaire.

In terms of the general landscape of artificial intelligence in the ICT-sector, the most notable finding is that companies in the sector mainly use machine learning as AI-technology for their businesses and provide predominantly AI-enabled services. In contrast, the research expected at least some organizations to apply AI to tangible products. This was not the case. For example, many of the companies described that clients have unused data, the interviewed ICT-companies then help them to make sense of their data. Another notable finding is that large multinational companies mostly develop their AI technologies in house, whereas SMEs often also use tools, plugins, and platforms such as that of IBM's Watson or Microsoft Azure. This has implications on the sphere of influence SMEs have with regards to ethics, specifically in the field of privacy and transparency because other parties could have access to data of these organizations.

The most pressing ethical challenges with AI that were found during the interviews are fairness and explainability. Challenges with fairness mainly related to the mitigation and prevention of unwanted biases. The absence of unwanted biases will be adopted to explain fairness in the questionnaire. On the other side, the key difficulty with explainability was that companies found it hard to explain a complex system or decision and to determine the level of understanding of a target audience that is required. This indicates that explainability is currently the key challenge, hence the survey will focus on this concept instead of a broader definition of transparency. Safety and privacy were also named as important ethical challenges for companies in the sector. Safety is mostly about preventing society from harm and is a difficult challenge for companies because of its unexpectedness and unpredictability. Lastly, as was to be expected, privacy is an important challenge for companies in the industry. Privacy considerations mainly arose in the data collection phase. Although this is one of the most challenges that companies most frequently think about, the GDPR provides a clear framework for organizations. Companies often limited their privacy practices to comply with the GDPR. Because privacy is already addressed extensively in NLdigital's code of ethics, this concept will not form a major building block for the survey.

Then, mechanisms that companies have in place for dealing with the abovementioned ethical challenges were discussed. In order to avoid unwanted biases in data, large multinational companies use fairness kits that check for fair distributions among sub-populations. SMEs do not have such tools and therefore either don't analyze it or check it manually. Another way in which fairness can be promoted is by having a diverse team throughout the machine learning life cycle. With regard to explainability, organizations try to build the ability to explain to users what is happening in their software tools. Another way is by actively providing information to their users and educating clients and explaining their model in a continuous client – company interaction. A

notable result was that for safety very few mechanisms are currently used. Organizations try to build as robust systems as possible by for example having a human check build into the loop, but do not have scenarios for the unforeseen impact that an AI-system might have. Organizations indicated that building a safety net is something they should do in the future, but currently, they are not doing so. These organizations do also not have an ethics board, either because they deem unnecessary or because they cannot afford it.

In the next step, internal mechanisms for increasing ethical behavior were discussed. A table of 16 key internal mechanisms and activities (Table 14) that organizations use and do to increase ethical behavior internally resulted from the interviews. This list of mechanisms will be used in the survey to evaluate which of these internal mechanisms are predominantly used in the industry, and what the distribution of these is among large organizations and SMEs. This is interesting because SMEs indicated to have significantly less of such mechanisms in place. This became particularly clear when organizations were inquired about the pillars of ethical governance that underlie their organization. Where large multinationals possessed nearly all of the pillars, SMEs did not possess a single one of them. The questionnaire will further evaluate the extent to which SMEs possess such pillars and internal mechanisms to generalize the interview results to a larger sample of the industry. A better understanding of the degree to which concrete measures and governance structures are used will help to identify opportunities to increase ethical behavior within SMEs.

Finally, feedback on NLdigital's code of ethics was collected. The first feedback of the interviewees was univocal positive about the eight principles that currently make up the code of ethics for the industry. Although the existing principles intuitively made sense and are all relevant to strive for according to participants, some of them were seen as fuzzy or limiting in guide-ability. Furthermore, participants indicated some important principles from their practice that they believe to be missing in the current code of ethics. Most frequently, the interviewees mentioned that there is no 'good' principle on fairness included in the code of ethics. The interviewees suggest that a principle addressing the practical challenge of analyzing and testing data with the goal to minimize bias, and a notion of what fairness means should be included in the code of ethics. Another suggested addition for fairness is a principle around people and data, namely, to actively work with a diverse group of people and stakeholders, including people from different backgrounds, ages, gender, etc., with the goal to incorporate ethical principles and inclusivity in the design of the product or service. Another principle that was seen as "fuzzy" was around the ease of explanation. Explainability is one of the key challenges with AI, and this needs to be better addressed. Organizations missed information about the target group of who to explain your system or model to, and the level of detail that should be included. Moreover, a principle or guideline that demands a feedback loop was seen as missing. The survey will build forward on these suggestions, and further investigate the need for better principles around the themes of fairness and explainability.

5. Questionnaire results

This chapter presents the results of the analysis of the data from the questionnaire. The questionnaire is attached in Appendix C. The electronic questionnaire recorded a total of 46 respondents from NLdigital (37) and the KNVI (9). In total the survey was sent out to over 750 people, this results in a response rate of around 6%. The resulting data from the questionnaire is analyzed using Microsoft Excel and IBM's SPSS.

First, section 5.1 delineates the process of data cleaning and provides a brief analysis of the characteristics of the data sample and respondents. Then, section 5.2 provides the results of the questionnaire with regards to the general AI-landscape and section 5.3 of the ethical challenges that the questionnaire found. Section 5.4 delineates the analysis of the ethical governance structure and mechanisms that companies have in place, and section 5.5 presents performance indicators and feedback on the current code of ethics. Finally, section 5.6 concludes the most notable results that result from the questionnaire.

5.1 Exploratory sample analysis

Before doing any meaningful analysis of the data to answer the research questions of this work, the data is first explored and treated. This section briefly summarizes the data cleaning process (section 5.1.1) and provides some descriptive statistics to understand the data and its distribution (section 5.1.2).

5.1.1 Data cleaning

The data from the questionnaire was registered in the 'Spotler' online survey environment. In order to make the data workable for further analysis, different treatment processes were executed. First, the data was exported in CSV format to be cleaned and treated using the Power Query editor of Microsoft Excel. The cleaning of the data is done in three steps. First, the original variables in the CSV file are renamed in order to increase readability and ease further analysis. Second, unnecessary and unwanted variables were deleted, and an extra indexing variable was created. Finally, test responses, incomplete responses, and outliers were deleted. This treatment process is delineated in more detail in Appendix D. After cleaning the data, the remaining dataset included 34 valid responses and 40 variables.

5.1.2 Exploratory descriptive statistics

To understand the diversity and distribution of the sample, exploratory data analysis is performed using IBM's SPSS. These analyses give an understanding of the businesses that we analyze throughout this chapter. First, in sub-section 5.1.2.1, the distribution of different sized companies is presented. Then, in sub-section 5.1.2.2 and sub-section 5.1.2.3 we look at the distribution of headquarters location and company footprint, respectively.

5.1.2.1 *Company size*

An important element of our hypothesis is that SMEs were underrepresented in the construction of the code of ethics and that the code of ethics may as a result not be as useful for them as for large organizations. Because we will look at differences between large organizations and SMEs throughout some parts of the rest of this analysis, it is important to understand the distribution of

these organizations among the respondents. This distribution is visualized in Figure 4. As the large organizations are commonly the more active ones within the NLdigital community, it is a favorable outcome that most of the respondents come from SMEs.

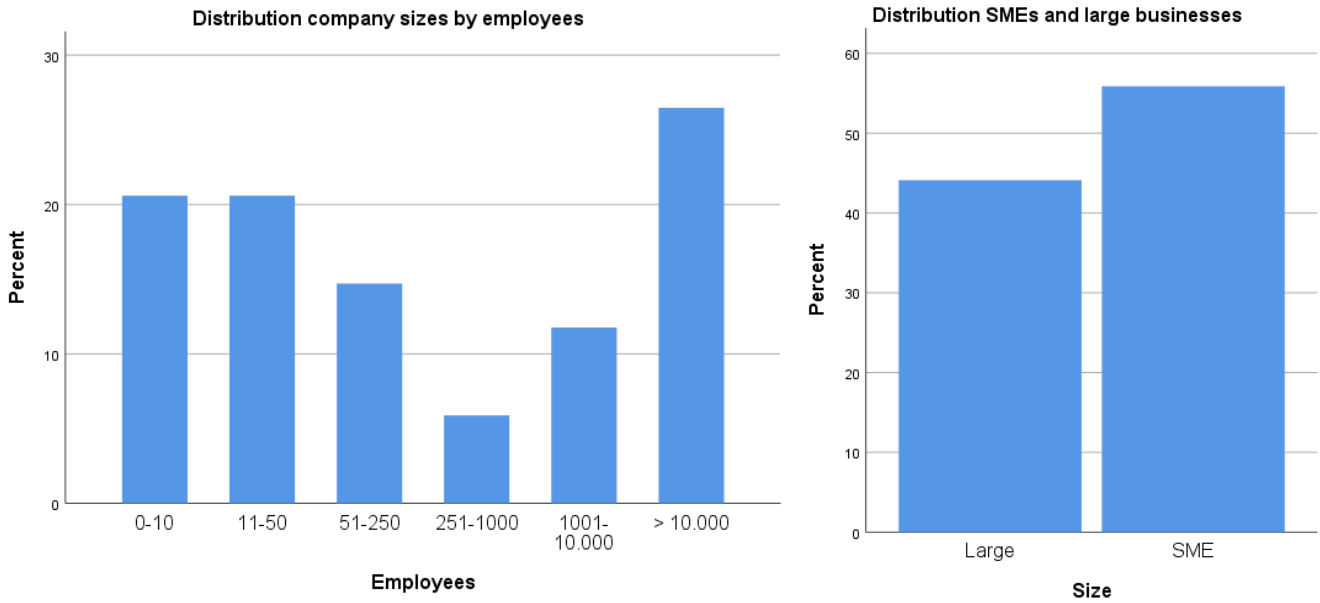


Figure 4: Distributions of respondent's organization by employees and company size

5.1.2.2 Headquarters of the organization

The majority of the company headquarters (74%) of the respondents are located in the Netherlands. It is no irregularity that all SMEs are headquartered in the Netherlands considering that this research has focused on the Dutch ICT-sector. The majority of the large international organizations are headquartered in the United States. Figure 5 shows the distribution of company headquarters within the sample for the different sized businesses.

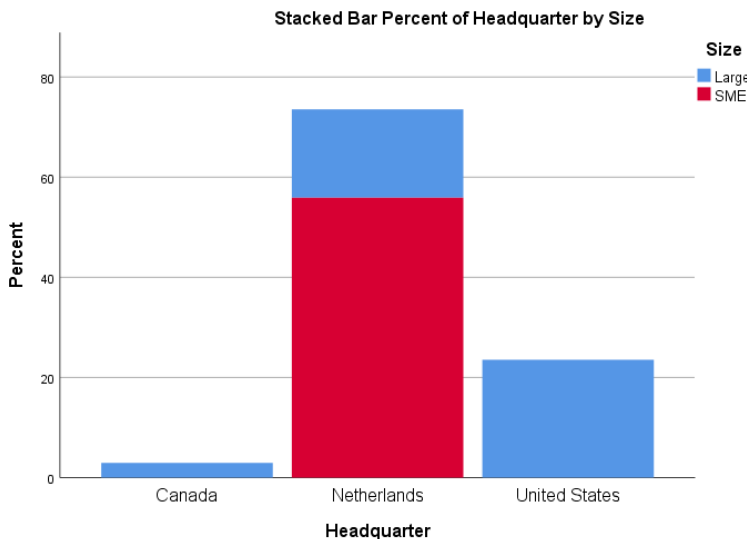


Figure 5: Company headquarters by size

5.1.2.3 Footprint of the organizations

Finally, the distribution of the footprint of the organizations in the sample is analyzed by company size (Figure 6). Again, we see a similar trend as with the company headquarters, i.e., SMEs have a more local character than the large organizations. Although there are some large organizations that are only active in the Dutch market, the vast majority has a global footprint.

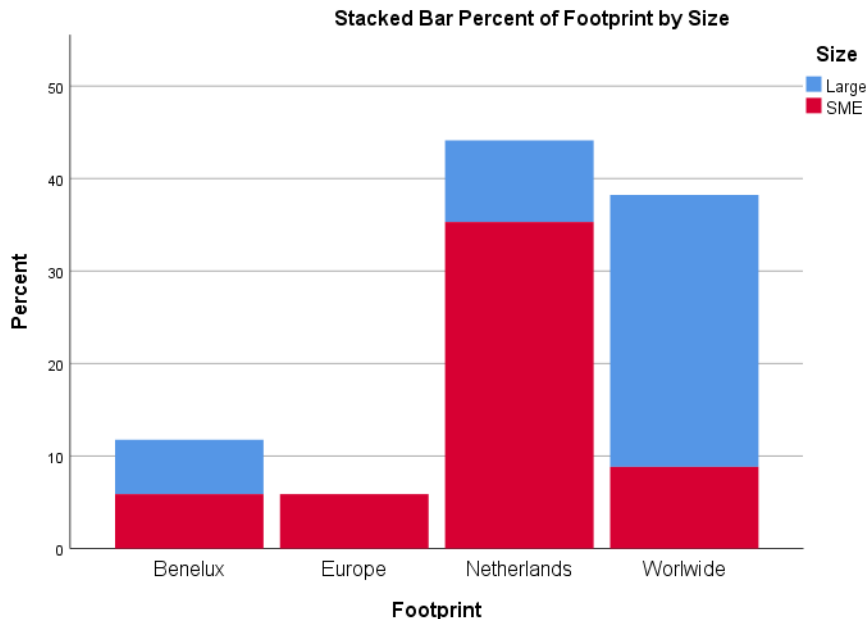


Figure 6: Company footprint by size

5.2 The landscape of artificial intelligence

For the analysis of the questionnaire results, the same structure as for the interview results is followed. This is because both methods aim to reinforce the conclusions that can be drawn from them. In this section, we briefly lay out the general findings of the artificial intelligence landscape within the ICT-sector in the Netherlands. Any data and graphs provided below are new numbers that were previously unknown about the ICT-sector in the Netherlands.

The interviewees for the semi-structured interviews were filtered on working with artificial intelligence to extract as much information about their work with AI-systems as possible. For the questionnaire, however, we provided equal opportunities to fill in the questionnaire for businesses that are working with AI and businesses that are not (yet) working with AI. This allows this work to show the distribution of businesses that use or don't use AI (Figure 7). The survey found that 79% of ICT-businesses active in the Dutch ICT-sector use AI for products, services, or internal processes. All respondents from large organizations indicated to use a form of AI. Of SMEs on the other hand, only 63% indicated to currently use AI-systems for products, services, or internal processes. In section 5.2.1 we briefly further discuss the businesses that are not yet working with AI and in section 5.2.2 we look at the purpose, autonomy, and stage of development of businesses that are working with AI.

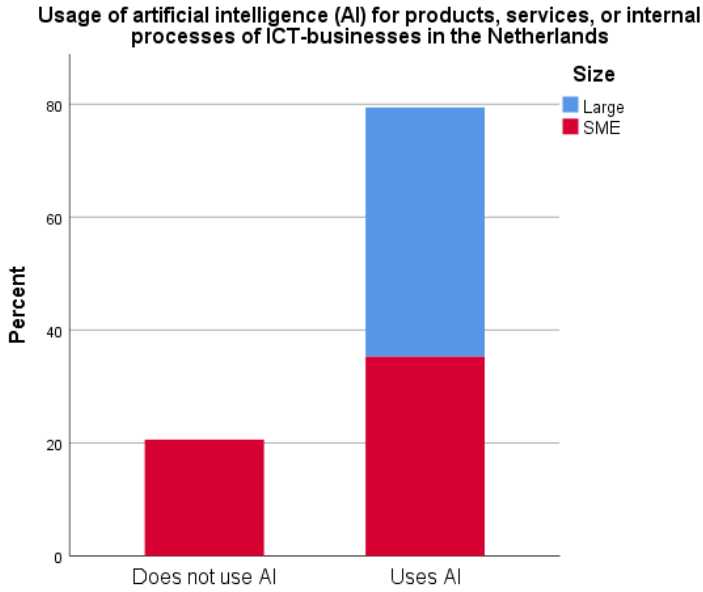


Figure 7: Usage of AI within ICT-businesses in the Netherlands

5.2.1 Businesses without AI-systems for products, services, or internal processes

Of the businesses that do not use AI yet, 57% indicated that they are planning to start using AI-systems. However, note that this number is based on a small sample size, n=7, because the majority of the sample did indicate to use a form of artificial intelligence for their products, processes or services. Apart from looking at the ethical challenges that businesses encounter in their use with AI, the questionnaire also aimed to find if ethical challenges might refrain businesses from using AI-systems. For reasons to not use AI-systems, respondents mostly indicated that 1) it is not useful or provides no additional value for their processes, products, or services (57%), and 2) they don't have the resources to do so (71%). Only one respondent indicated that ethical challenges and potential negative consequences are holding them back from using AI-systems.

5.2.2 Businesses using AI-systems for products, services, or internal processes

Businesses that indicated to use AI-systems can be divided into different stages of the development process. The majority of the organizations have started with using AI-systems for some processes, products or services (Figure 8). However, large organizations are in a further stage than SMEs. This could provide additional evidence for the interview finding and initial assumption that SMEs are currently more focused on developing artificial intelligence systems from a technical perspective than already worrying about every ethical aspect.

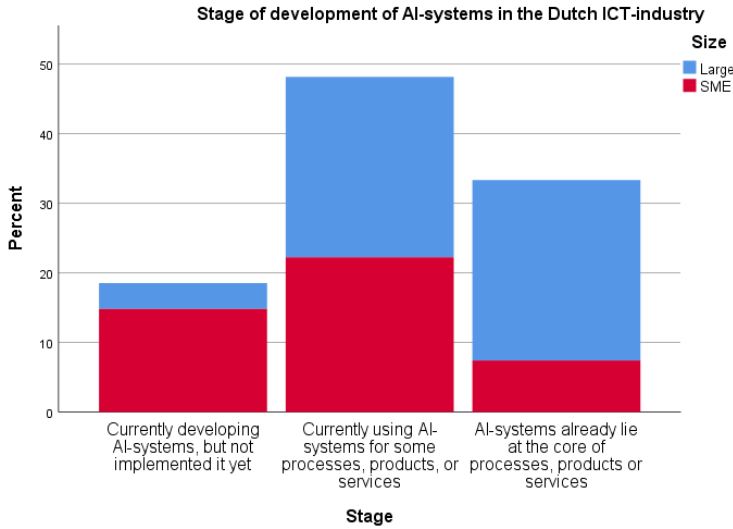


Figure 8: Stage of development of AI-systems in the Dutch ICT-industry

In terms of how ICT-businesses develop AI-systems a distinction can be made between three groups (Figure 9), i.e., organizations that completely develop their own AI-systems (33%), organizations that develop their own AI-systems with the help of services of others (56%), and organizations that do not develop any AI-systems but only use services of others (11%). Large multinationals mostly develop their own systems whereas SMEs predominantly make use of services of others. This has some consequences for the sphere of influence that SMEs have on their ethical practices. For example, in the development of the AI-system, many SMEs will use the services of others, predominantly the large ICT-organizations, to make their products or services functional. This can influence, for example, the SME's privacy practices and the extent to which they can built-in technical solutions to resolve ethical challenges. For privacy, this means that the data that they use (often from their customers) will sometimes flow to the larger organization that helps build their AI-systems. While for resolving some ethical challenges by building technical solutions into the AI-system, the SME can be dependent on the technical features, software, and services that the large organization offers.

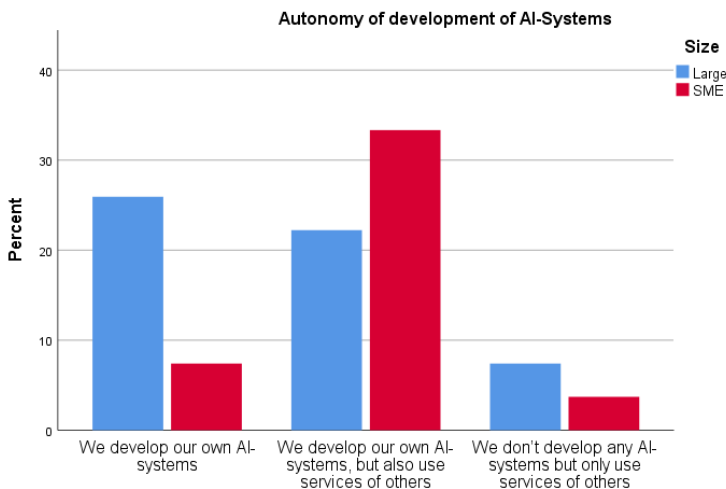


Figure 9: Autonomy of development of AI-systems

5.3 Pressing ethical challenges

The semi-structured interviews were the first methodology to inquire into the most pressing challenges that businesses within the industry encounter in the development and application of AI-systems. The interviews laid the qualitative basis for the survey item (Appendix C) by defining the ethical challenges as they were interpreted by the industry, together with the dominant aspects of the literature. For the interviews, this resulted in fairness, explainability, safety, and privacy being the most pressing challenges. In the survey, these challenges were defined as the following:

- Fairness: prevention of biases that cause 'unfair' outcomes
- Explainability: explaining predictions, recommendations, or decisions
- Safety and non-maleficence: preventing people from harm
- Privacy: collecting and protecting people's data and privacy in a responsible way
- Responsibility: who is responsible and accountable for the consequences of AI-systems
- Transparency: acts of communication about the AI-system

Besides these options, respondents were given the opportunity to list other challenges or definitions or indicate that they are not experiencing any pressing ethical challenges. Respondents were only able to list a maximum of two most pressing ethical challenges in order to provide a clear and decisive ranking of importance for the challenges.

Figure 10 provides an overview of the frequencies of the response to the different ethical challenges. Most of the respondents took the opportunity to list two ethical challenges as most pressing (181% responses). In line with the interview results, explainability is considered the most pressing challenge for businesses. Of all the businesses, 56% indicated that explaining predictions, recommendations, or decisions is the most pressing ethical challenge with AI for their business. Privacy (37%) and fairness (33%) are also still considered as important challenges for businesses in the industry as they make up the second and third most pressing challenge, respectively. An interesting new finding from the questionnaire is that responsibility is considered a pressing challenge by almost a quarter of the industry, especially among SMEs it is considered a significant challenge. A possible explanation for this is that companies might have refrained from mentioning responsibility as a pressing challenge during the interviews out of fear for damaging the reputation of their business.

If we further look at the differences in challenges between large organizations and SMEs it can be seen that large organizations are struggling more with the safety aspect. This could mean that large organizations are more aware of this challenge, that this challenge starts to play a role in the later development stage, or that they have insufficient means, tools, mechanisms, or knowledge to deal with this challenge. Nonetheless, it reinforces the question of why safety is not yet represented in NLdigital's code of ethics.

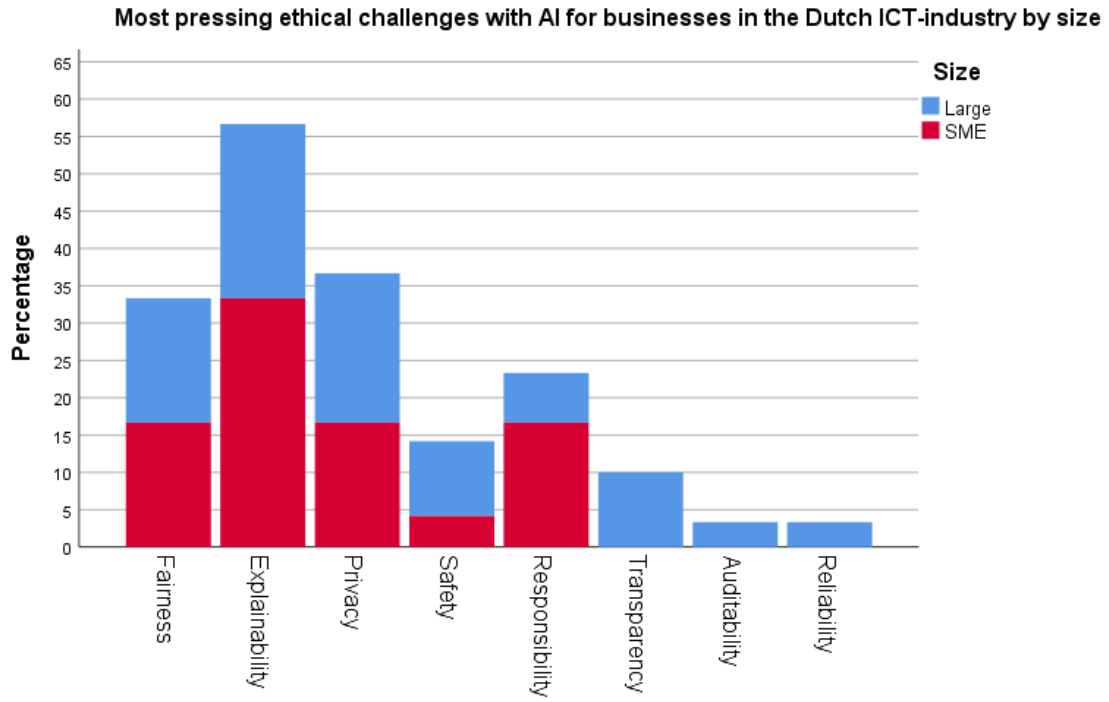


Figure 10: Most pressing ethical challenges with AI for businesses in the Dutch ICT-industry by size from the questionnaire

Sub-research question 4: Which ethical challenges are most pressing to Dutch ICT-businesses in the development and application of AI-systems?

Partial answer (part 2): In section 4.2.1 we answered the first part of this research question and found that businesses listed explainability, fairness, safety, and privacy as their most pressing challenges in the development and application of AI-systems. Concretely, these challenges were described to be:

- Prevention of biases that cause ‘unfair’ outcomes (fairness)
- Explaining predictions, recommendations, or decisions (explainability)
- Preventing people from harm (safety and non-maleficence)
- Collecting and protecting people’s data and privacy in a responsible way (privacy)

The questionnaire further investigated the importance and gravity of these challenges for businesses in the Dutch ICT-industry. This method also found explainability, fairness, and privacy as the most pressing challenges in the industry. The questionnaire thus reinforces and validates the interview results that these are challenges that both large businesses and SMEs struggle with. Interestingly, the distribution of these challenges among SMEs and large organizations is reasonably equal. This indicates that not only SMEs, but also large companies still struggle with dealing with fairness, explainability, and privacy challenges. From the interviews, however, we know that businesses have strict guidelines through the GDPR. The indication of privacy as one of the most pressing challenges is thus likely to be a result of privacy being a continuous topic of consideration in the development and application of AI-systems rather than a lack of knowledge, measures, tools, and mechanisms of how to deal with it. Although considerably less pressing, safety, and responsibility, were still found to be significant ethical challenges for organizations as well. Table 16 summarizes the key ethical challenges that are most pressing to Dutch ICT-businesses in the development and application of AI-systems.

Table 16: Most pressing ethical challenges with AI for businesses in the Dutch ICT-industry

Challenge	Interview count (n=6)	Survey percentage (n=27)
1. Explainability	4	56 %
2. Fairness	4	33 %
3. Privacy	2	37 %
4. Safety	2	15 %
5. Responsibility	-	22 %

5.4 Ethical governance

Building on the results of the interviews (section 4.3), this section discusses the mechanisms and activities that organizations have in place to encourage, enforce, and assist ethical behavior with artificial intelligence for their employees. The first sub-section (5.4.1) delineates the results of testing the internal mechanisms for increasing ethical behavior that resulted from the interviews for a larger sample. Subsequently, sub-section 5.4.2 displays the results of the inquiry into the pillars for good ethical governance of AI that are present in businesses in the industry (Winfield & Jirotko, 2018).

5.4.1 Internal mechanisms for increasing ethical behavior

During the semi-structured interviews, interviewees named various mechanisms and measures that they use to encourage, enforce, and assist ethical behavior with artificial intelligence for their employees and their business. Table 14 in section 4.3.2 listed and explained these different mechanisms and measures. In the questionnaire, respondents were given the opportunity to select multiple internal mechanisms and measures out of this initial list. Table 17 below provides an overview of the overall results, where the most occurring mechanisms are listed on top.

Table 17: Frequency of internal mechanisms to encourage ethical behavior

		Responses		Percent of Cases
		N	Percent	
Mechanisms ^a	An open culture to discuss AI and ethics	20	11.3%	74.1%
	Collaborate actively with the client and/or customer	17	9.6%	63.0%
	A feedback mechanism for users	14	7.9%	51.9%
	Strong ethical leadership in the organization	14	7.9%	51.9%
	Peer review each other's work	14	7.9%	51.9%
	AI and ethics training	12	6.8%	44.4%
	Publish publications, videos, and documentaries	11	6.2%	40.7%
	Seek help at communities and colleagues for ethical challenges	11	6.2%	40.7%
	A written down set of rules or principles by which everyone complies	11	6.2%	40.7%
	A feedback loop with ethics checks throughout the life cycle of projects	9	5.1%	33.3%
	Punish or fire people that behave unethically	7	4.0%	25.9%
	A secure channel where employees can expound their moral concerns about the company's practices	5	2.8%	18.5%
	The central storage system in which all AI models and data are stored	3	1.7%	11.1%
	Other: Ethical review board	2	1.1%	7.4%
Total	177	100.0%	655.6%	

a. Dichotomy group tabulated at value 1.

A total of five mechanisms are present within more than half of the businesses, with 'An open culture to discuss AI and ethics' being responded the most. These results suggest that the majority of companies are actively trying to stimulate ethical behavior within their AI-practices. However, notable is that almost all five of these mechanisms (except for providing a feedback mechanism for users) are relatively unmeasurable and 'intangible' mechanisms. They have a larger component of personal interpretation and experience than the more tangible ones such as training, publications, written down ethics principles, etc.

The preference for unmeasurable mechanisms is especially prevalent when we distinguish the internal mechanisms by company size (Figure 11). Where large companies seem to have a more

balanced set of mechanisms and a much higher presence in general, SMEs mainly possess the more unmeasurable mechanisms that are subject to personal interpretation. The results indicate that only a small fraction of the SMEs provide AI-Ethics training, publish documents, punish unethical behavior, have ethics communities, work in a feedback loop with ethics checks, have channels for reporting misconduct, and publish a set of ethics principles. The fact that some of these mechanisms are present in many large organizations highlights a significant gap. Especially with regards to the following mechanisms:

- AI-Ethics training
- Publishing of publications, videos, or documentaries
- Punishing unethical behavior
- Ethics communities
- A feedback loop with ethics checks throughout the life cycle of projects
- Secure channel for reporting moral misconduct
- A set of ethics principles.

Percent of businesses with different internal mechanisms to encourage ethical behavior present

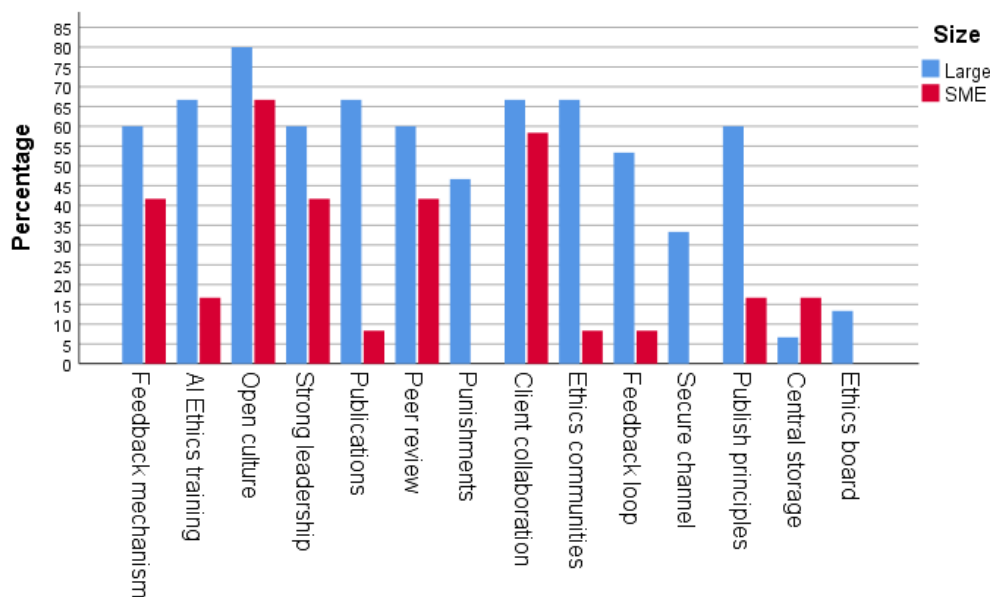


Figure 11: Internal ethical mechanisms by company size

5.4.2 Pillars of ethical governance

Respondents were asked to indicate which pillars for good ethical governance are present in their organization. These pillars follow the framework that was proposed by Winfield and Jirotko (2018) and previously described in section 3.2.3.2. The results of the presence of these pillars among ICT-businesses in the survey is shown below in Figure 12.

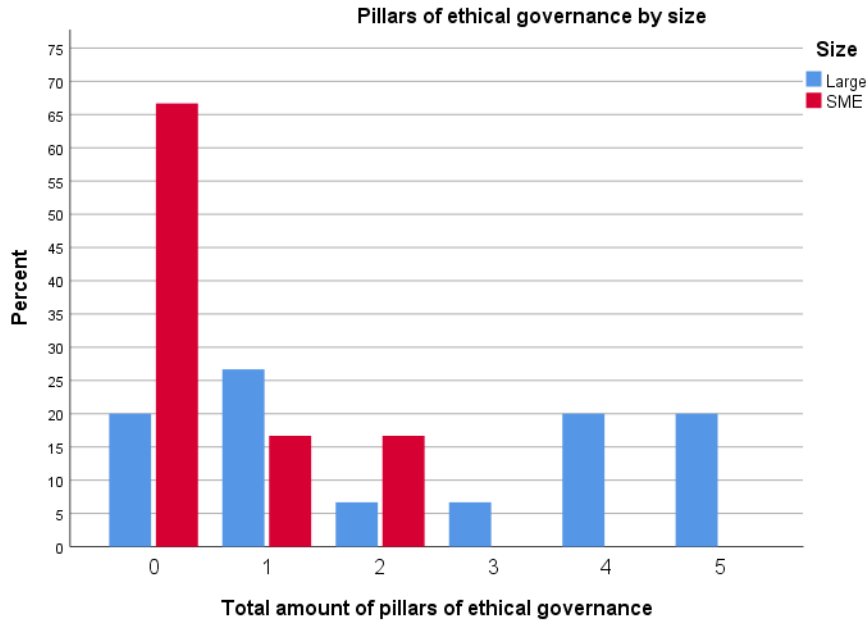


Figure 12: Number of pillars of ethical governance in organizations by size

The interview results suggested a clear trend of large companies having vastly more ethical pillars on which they built their ethics practices than the SMEs in the industry. The survey results further reinforce this finding and increase the reliability of both methods. For example, none of the SMEs had more than two pillars in place, whereas for large organizations this is almost 47% of the total number of large organizations. Furthermore, 67% of all SMEs possess zero of Winfield & Jirotká's pillars of ethical governance. This is a striking number considering these pillars are deemed necessary for solid ethical governance of artificial intelligence systems within organizations.

When we look at the presence of each of the specific pillars in Figure 13, or one of the parts that make up a pillar (in the case of pillar 1 and 3), some other findings can be established. First of all, it stands out that none of the SMEs publishes a code of ethics, while on the other hand over half of the large organizations do publish one. This is an important finding because literature has suggested its importance. This research recognizes that it is understandable that SMEs have fewer pillars present because of the smaller resources and often the earlier stage that they're in, but that this is not something to strive for. Additional mechanisms, tools, and structures will need to be developed by and for SMEs to ensure solid ethical governance of AI in the sector. The lack of such pillars within SMEs currently, especially the lack of a code of ethics, strongly supports the value the code of ethics from NLdigital can have for them.

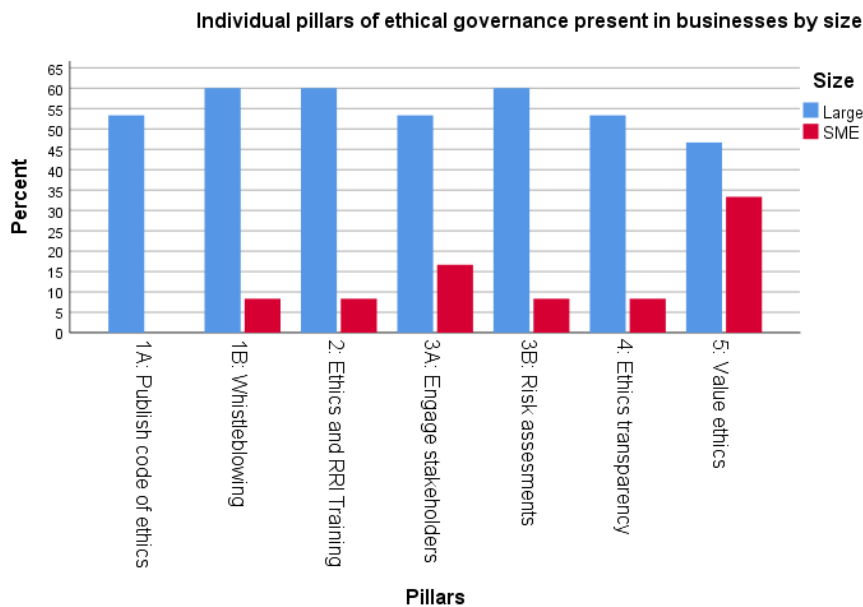


Figure 13: Presence of individual (parts) of pillars of ethical governance

Sub-research question 5: What concrete measures do Dutch ICT-businesses currently use to deal with the ethical aspects of developing and applying AI-systems?

Partial answer (part 2): In part one of the answer to sub-question 5 (section 4.3), we discussed specific tools and measures that interviewed businesses indicated to use to deal with individual challenges such as for explainability (i.e., educating clients, build explainability into software) and fairness (i.e. fairness software kits, working in diverse teams). For the second part, the questionnaire tested to what extent ICT-businesses use different concrete internal measures to encourage and facilitate ethical behavior include. The results indicate that a total of five measures are present within more than half of the businesses. In order of importance these are:

1. An open culture to discuss AI and ethics
 2. Collaborate actively with the client and/or customer
 3. A feedback mechanism for users
- Strong ethical leadership in the organization
Peer review each other's work

More concrete mechanisms are used in large organizations, these include AI-Ethics training, publish documents, punish unethical behavior, have ethics communities, work in a feedback loop with ethics checks, have channels for reporting misconduct, and publish a set of ethics principles. The fact that some of these mechanisms are present in many large organizations highlights a significant gap in ethical practices with SMEs. An inquiry into the pillars of ethical governance also discovered that larger organizations commonly have more ethical governance measures in place, e.g. publish a code of ethics, have a whistleblower channel, provide ethics and RI training, undertake ethical risk assessments, and engage a wide range of stakeholders. SMEs are lagging behind vastly in this respect, with nearly 67% having zero pillars of good ethical governance in place.

5.5 Code of ethics

Several questions in the survey were dedicated to mapping out the importance, adherence, and adoption of the current code of ethics. The goal of this section is to further identify opportunities for an improved code of ethics. Section 5.5.1 presents the adoption of the code of ethics, section 5.5.2 displays the results of the questions on the adherence of the code, and finally, section 5.5.3 considers the results of the feedback statements that respondents were asked to answer.

5.5.1 Adoption of the code

With regard to the familiarity with the current code of ethics, only 33% of the industry is familiar with it. Of this small portion, the majority are large organizations (of which some have most likely participated in the creation of the code). As can be seen in Figure 14, only 17% of the SMEs that responded to the survey indicated that they are familiar with the code of ethics. Because of the limited amount of data points left after removing those respondents that are unfamiliar with the code, no meaningful conclusions could be drawn about the frequency of usage of the code. The low familiarity rate also indicates that apart from improved ethical principles, communication of the code should also be improved.



Figure 14: Familiarity of businesses with NLdigital's code of ethics

5.5.2 Adherence to the code

Respondents were asked to express the extent to which they carry out the activities that are prescribed in the current code of ethics. 'Principle 1' was excluded from the analysis because it does not contain a tangible action.

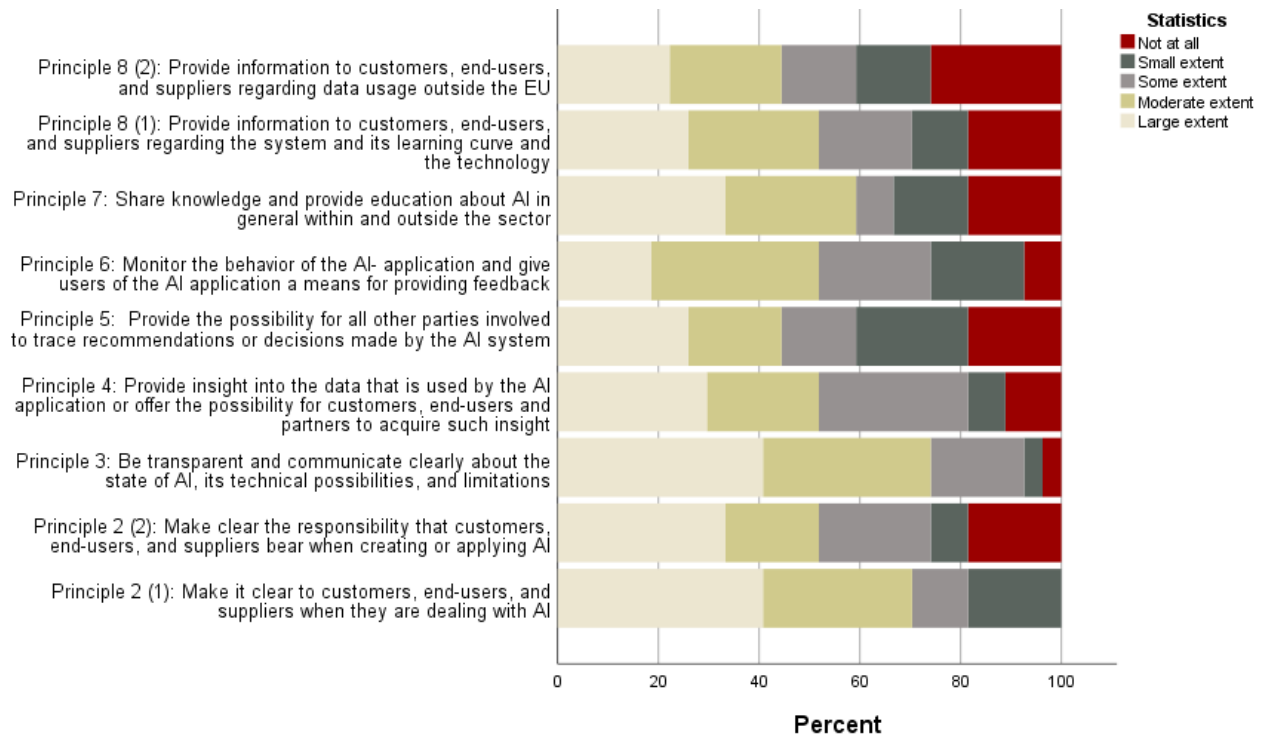


Figure 15: Extent to which organizations follow the principles of the code of ethics

As can be seen in Figure 15, the vast majority of the principles are carried out to some extent. This indicates 1) that the current principles are achievable for many of the businesses in the sector and 2) that companies are actively working on these 'ethical' activities. The most adhered to principles are:

- Make it clear to customers, end-users, and suppliers when they are dealing with AI (*Principle 2 (1)*)
- Be transparent and communicate clearly about the state of AI, its technical possibilities, and limitations (*Principle 3*)

Interestingly, one of the least adopted principles is the last part of principle 8, i.e., provide information to customers, end-users, and suppliers regarding data usage outside the EU. Under the adequacy decision of the EU, it is not necessary to inform or ask the consent of customers if data is flowing to countries that are recognized by the European Union, e.g., the United States and Israel. This provides further evidence for the finding that ICT-businesses currently limit their practices mostly to what the GDPR demands. Another scarcely adhered to principle is principle 5: 'provide the possibility for all other parties involved to trace recommendations or decisions made by the AI system'.

5.5.3 Feedback statements

Finally, in the last part of the survey, businesses were asked to indicate to what extent they agree with the main feedback from the interviews. Figure 16 shows the results of the statements on a Likert scale for both SMEs and large organizations. First, both SMEs and large organizations agree that a code of ethics is necessary to encourage ethical behavior with artificial intelligence.

This result lays bare a gap that exists between what companies (specifically SMEs) believe is necessary to guide ethical behavior and what they are doing in practice. In section 5.4, SMEs indicated earlier to not publish a code of ethics. A second observation we can make based on the results in Figure 16 is that both SMEs and large organizations believe that explainability, fairness, and safety should be included in a code of ethics for artificial intelligence. This underlines the need for including a safety principle in the code and analyzing if the current principles on fairness and explainability are adequate enough. Finally, contrary to our assumption that SMEs would imply that more ethics mechanisms should be made available to their organization, most SMEs rated the statement neutral. This is a surprising outcome considering that our other results indicate that SMEs are lacking a very large number of mechanisms, measures, and pillars to encourage and govern AI ethics.

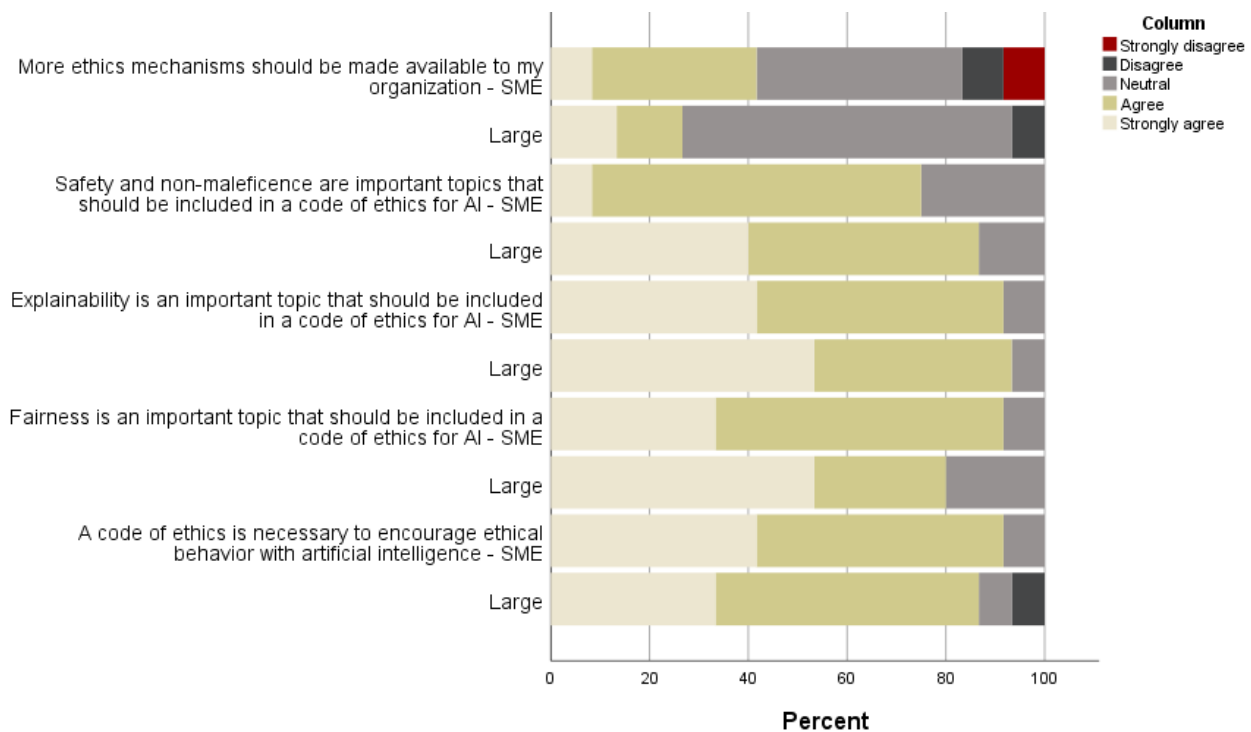


Figure 16: Feedback on code of ethics statements measured on a Likert scale

5.6 Conclusion

This chapter presented the results of the survey that was sent out to NLdigital’s members and some members of the KNVI. The type of survey that was chosen for this research is an electronic questionnaire. This is generally considered a valuable data collection method for collecting descriptive and exploratory information such as is the case in this research (Sekaran & Bougie, 2016, p. 147). The goal of the questionnaire was to map out the general AI-landscape in the Dutch ICT-industry, determine the importance of the different ethical challenges with AI, and map out the mechanisms, pillars, and measures that businesses use to govern, encourage or enforce ethical behavior. The nature of the questionnaire is thus descriptive, and its primary goal was to provide a descriptive analysis with regards to the abovementioned concepts.

In terms of the general landscape of artificial intelligence in the ICT-sector, the results indicate that 79% of the ICT-businesses active in the Dutch sector use AI for their products, services, or internal processes. When distributed over the two categories (i.e., large organizations and SMEs), it was revealed that all large businesses that responded use some form of AI, while for SMEs this is a much lower number (63%). Although more than half of the organizations that do not use AI yet indicated that they are planning to do so, the main reasons not to work with AI technology are 1) it is not useful or provides no additional value for their processes, products, or services, and 2) they do not have the resources to develop and employ such technologies. For businesses that do actively work with AI-systems, the questionnaire looked at how such companies develop their systems. Overall, most companies in the industry develop their AI-systems partly by themselves but also with the help of some services of others. Large organizations, however, predominantly develop their own systems and are in a further stage of the AI development process than SMEs. This provides additional evidence for the interview finding and one of the initial assumptions that SMEs are currently more focused on developing artificial intelligence systems from a technical perspective than occupied with every ethical aspect. The fact that many SMEs will use services of large organizations in the development of their AI-systems raises the question if these large businesses should or could play a role in protecting and educating small and medium-sized enterprises, especially considering their often more advanced knowledge and resources on AI-ethics.

To prove the interview findings, the questionnaire also tested the most pressing ethical challenges with AI-systems for businesses in the ICT-sector. The results of both methods combined resulted in the answer to sub-research question 4 (section 5.3). In line with the interview results, explainability is considered the most pressing challenge for businesses. Of all the businesses, 56% indicated that explaining predictions, recommendations, or decisions is the most pressing ethical challenge with AI for their business. The results of the questionnaire also demonstrate fairness and privacy among the most pressing challenges in the industry. The questionnaire thus reinforces and validates the interview results that these are challenges that both large businesses and SMEs struggle with. Interestingly, the distribution of these challenges among SMEs and large organizations is reasonably equal. This indicates that not only SMEs, but also large companies still struggle with dealing with fairness, explainability, and privacy challenges. From the interviews, however, we know that businesses have strict guidelines through the GDPR. The indication of privacy as one of the most pressing challenges is thus likely to be a result of privacy being a continuous topic of consideration in the development and application of AI-systems rather than a lack of knowledge, measures, tools, and mechanisms of how to deal with it. If we further zoom in on the differences in challenges between large organizations and SMEs it can be concluded that large organizations view the safety aspect as more pressing. This could mean that large organizations are more aware of this challenge, that this challenge starts to play a role in the later development stage, or that they have insufficient means, tools, mechanisms, or knowledge to deal with this challenge. Nonetheless, it reinforces the question of why safety is not yet represented in NLdigital's code of ethics.

For dealing with these challenges and to enforce ethical behavior with AI within their business in general, ICT-organizations use different measures (also referred to as mechanisms throughout this work). The survey points out that out of the measures discovered during the interviews, a total

of five different measures are present within more than half of the businesses. The mechanism with the highest presence is an open culture to discuss AI and ethics. These results suggest that the majority of companies are actively trying to stimulate ethical behavior within their AI-practices. However, notable is that almost all five of these mechanisms (except for providing a feedback mechanism for users) are relatively unmeasurable and 'intangible' mechanisms. On the other hand, tangible and demonstrable measures that are less reliant on personal interpretation and opinion are still lacking in the majority of businesses, especially in SMEs. The analysis of pillars of ethical governance for companies in the industry showed that large companies have vastly more ethical pillars on which they built their ethics practices than SMEs. In total, 67% of all SMEs possess zero pillars and none of the SMEs had more than two out of the five pillars of ethical governance in place. This is an alarming number considering these pillars are deemed necessary for solid ethical governance of artificial intelligence systems within organizations. Yet, it is also unsurprising given the early stage they are in and the limited amount of resources they have relative to their larger counterparts. Nonetheless, the negative impact that SMEs could have considering the scalability of AI-technology remains. Therefore, we argue that mechanisms such as ethics and RRI training, whistleblower mechanisms, templates for risk assessments, and written down ethics principles should be developed for or by these organizations. Although the question for who this responsibility lies remains, the results provide strong motivation for the importance of NLdigital's overarching code of ethics for the ICT-industry.

Thus, good ethical code will have substantial value for the entire industry. Most businesses, and specifically most SMEs specified agree that a code of ethics is necessary to encourage ethical behavior with artificial intelligence. Supporting the academic literature and our analysis, they also stressed the need for having clear explainability, fairness, and safety principles in the code of ethics. Nonetheless, the results of the questionnaire also demonstrate that businesses currently are largely unfamiliar with the code of ethics. Apart from an improved code of ethics, more and better communication throughout all layers of the organizations within the community will be necessary for successful large-scale adoption of the code.

6. Conclusions

Following the literature review, the code of ethics analysis, semi-structured interviews, and the questionnaire, this chapter presents the conclusions of this research. These different methods that were used throughout this study are triangulated to arrive at the main findings in section 6.1. As previously discussed, this research and the used research methods are also bounded to some limitations which are discussed in section 6.2. In section 6.3 the implications of the main findings are discussed and a brief normative ethical reflection on the findings with respect to the code of ethics is presented. Subsequently, section 6.4 provides the recommendations that follow from this work both for the code of ethics and additional measures that are important for better ethical governance within ICT-organizations. Building forth on the scientific, industrial, and societal contributions of this work discussed in section 1.5, in section 6.5 we identify opportunities for future research. Finally, in section 6.6, the relevance to the 'Management of Technology' master program, of which this work is the dissertation, is outlined.

6.1 Main findings

This section discusses the main findings of this work. The literature review, code of ethics analysis, semi-structured interview results, and the questionnaire results are triangulated to arrive at the answers to the main research question.

6.1.1 Towards an improved code of ethics: answer to the research question

Bringing more artificial intelligence into practice in an ethically thoughtful way will offer the Netherlands many societal and economic benefits. The code of ethics from NLdigital is designed to guide organizations within the ICT-industry through the ethical challenges that AI-systems bring. The code is a living document that requires continuous improvement. The first version of the code was made with little knowledge of the concrete ethical challenges in the development and application of AI-systems in the industry, especially from SMEs. Rather, it was a collaboration of a handful of experts and large corporations. Contributing to further inclusion of data from SMEs and the process of improving the current code of ethics, the goal of this work was to **establish ethical challenges that are currently missing or not addressed in-depth in the code of ethics from NLdigital and are fundamental for guiding responsible research and development in the use and application of AI-systems for the Dutch ICT-industry**. To gather the necessary empirical evidence to answer this research question, a mixed-method approach was found to be most suitable, i.e., the exploratory sequential method. This approach consisted of an extensive literature review, analysis of NLdigital's guideline, semi-structured interviews, and an online questionnaire.

The literature review found 1) fairness, 2) responsibility, accountability, and trust, 3) privacy, 4) safety and non-maleficence, 5) transparency and explainability to be the most important ethical aspects in the research, development, and application of AI-systems for businesses. Then, an analysis of NLdigital's code of ethics and a comparison with other codes lead to the identification of three shortcomings, i.e., an unpractical principle of fairness, an unclear concept and structure of explainability, and the absence of a principle for safety. Subsequently, the interviews lead us to identify that fairness, explainability, and safety, form the basis of the most pressing challenges

for businesses in the sector. The results of the online questionnaire reinforced these results strongly, especially the importance of explainability and fairness. Lastly, in the interviews and questionnaire businesses indicated that it is precisely these aspects (i.e., fairness, explainability, and safety) that the code of ethics fails to address adequately for them. This research, therefore, arrives at the conclusion that explainability, fairness, and safety are currently addressed insufficiently in the code of ethics from NLdigital and are fundamental for guiding responsible research and development in the use and application of AI-systems for the Dutch ICT-industry. The sub-sections below motivate each of the aspects in more detail.

6.1.1.1 *Fairness*

Fairness is widely considered as one of the most important ethical principles in western societies. In section 3.2.2.3 we showed that the right to equality and non-discrimination forms the first article of the Dutch constitution, as well as a fundamental element of international human rights law such as the Universal Declaration of Human Rights (UNG Assembly, 1948). Fairness is also part of Ross's prima facie duties (Ross, 2002). For AI-systems specifically, fairness offers some novel challenges. Particularly, algorithmic fairness is an important topic discussed in the field of artificial intelligence because algorithms may act unfairly towards minority groups without any form of human intervention or control. The way in which biased data causes biased recommendations or decisions by AI systems is because machine learning systems are trained with the biased data, and will reproduce or reinforce these biases (Chouldechova & Roth, 2018). Algorithms can then, for example, treat groups differently with respect to gender, race, ethnicity, and disabilities (Zheng et al., 2018). Through examples of Amazon and Apple (section 3.2.2.3), it became clear that building 'unfair' AI-systems can not only lead to unfavorable results for individuals and society, but also for the businesses that build them. It is thus in everybody's interest to build fair systems.

Simultaneously, fairness was found to be one of the most pressing challenges for both large organizations and SMEs in the Dutch ICT-industry. In the interviews, four out of six interviewees indicated that fairness is one of their most pressing challenges. While in the questionnaire 33% expressed fairness as the most pressing challenge in the development and application of AI-systems. Challenges with fairness mainly related to the mitigation and prevention of unwanted biases (section 4.2.1.1). These biases may result in discrimination based on, for example, gender and ethnicity, as well as an unfair allocation of services and their corresponding prices.

Being an important challenge for the reasons found in literature, and in practicality in the ICT-sector, one would expect fairness to be strongly represented and formulated in the current code of ethics from NLdigital. However, our analysis of fairness in the code of ethics concludes that it is not addressed in a meaningful way for organizations in the industry. For fairness, the principle reads: "the member company is aware of and transparent regarding the state of the technique for AI and its technical possibilities but is also aware of the technical limitations of applications and the necessity to communicate clearly regarding this. This leads (among other things) to minimizing undesired "bias" and to promoting inclusive representation" (NLdigital, 2019, p. 1). This principle is formulated ambiguously, reads a dubious causality and provides minimal 'guidance' to organizations within the industry. It provides no norm for what fair is, what is expected of organizations with regards to fairness, and how fair AI-systems can be achieved.

Moreover, in feedback on the code of ethics during the interviews, interviewees also frequently mentioned that there is no 'good' principle on fairness included in the code of ethics. Although the word bias is mentioned once in the code, the majority of the interviewees indicated that the principle was 'fuzzy and insufficient to provide any value for their practice. The principle only says to be transparent but fails to address any direction about what you can and cannot do. It fails to address here the practical challenges in the work of a data scientist. Interviewees suggested that a principle addressing the practical challenge of analyzing and testing data with the goal to minimize bias, and a notion of what fairness means should be included in the code of ethics.

In summary, fairness is a fundamental ethical principle for guiding responsible research in the application and development of AI-systems, is one of the most pressing ethical challenges for businesses in the ICT-industry, and is according to our analysis as well as to feedback from the industry, insufficiently addressed in the current code of ethics.

6.1.1.2 Explainability

Transparency and explainability are considered as other very important ethical aspects for AI among academia. In section 3.2.2.7 we showed that explainability is important because it forms the groundwork for many of the other ethical challenges of AI in businesses. If an undesirable or harmful event occurs, explainability is needed so that the decision-making and circumstances can be understood and used as evidence of how something happened (Bryson & Winfield, 2017). In that respect, it is also needed to assess the responsibility and accountability of different actors. Section 3.2.2.7 showed the same line of reasoning for fairness, i.e., a clear understanding of the population and the logical steps an algorithm goes through are necessary to determine the fairness of a decision. Moreover, a lack of transparency and explainability forms a major barrier for the widespread acceptance and adoption of AI by both professionals and society (Owotoki & Mayer-Lindenberg, 2007). The reasoning process is often a vague black box that provides no insight for users and other stakeholders as to how decisions are made. This causes a large sense of distrust among these same groups. Explaining AI-systems in a helpful manner will lay the groundwork for many of the other ethical challenges and create a sense of trust with the general public that is necessary for the more widespread adoption of AI.

Nonetheless, explainability was found to be the most pressing challenge for both large organizations and SMEs in the Dutch ICT-industry. In the interviews, four out of six interviewees indicated that explainability is one of their most pressing challenges. While in the questionnaire 56% expressed explainability as the most pressing challenge in the development and application of AI-systems. All interviewees specifically named explainability rather than the wider concept of transparency. Businesses revealed that the key difficulty of explaining predictions, recommendations, or decisions is the complexity of the models and decision-making process. Furthermore, ICT-organizations are finding it difficult to find the right level of abstraction in explainability to deal with the different levels of understanding of clients, customers, and consumers.

Being an important challenge for the reasons found in literature, and a very pressing one for businesses in the ICT-sector, one would expect explainability to be strongly represented and formulated in the current code of ethics from NLdigital. However, our analysis of explainability in the code of ethics concludes that it is not addressed in a meaningful way for organizations in the

industry. Even though explainability is briefly mentioned in a sub-sentence in the first principle, it does so in an obscure way. Even if one finds the word 'ease of explanation' in the first principle, it is still unclear what is expected from an organization. With regards to explainability, the principle states: "the member company is aware of and anticipates the influence that implementation of AI can have on public values such as the principle of ease of explanation" (NLdigital, 2019, p. 1). The analysis in section 3.3.2 concludes that this does not cover off explainability sufficiently. It is a value-based statement that asks companies to be aware of explainability but offers little practical assistance for daily ethical decision making.

In addition, feedback from organizations on the code of ethics during the interviews declared the current principle of explanation in the code of ethics to be 'fuzzy' as well. Particularly, businesses miss information about the target group of who to explain your system or model to and the level of detail that should be included. Finding the right balance between interpretability and level of detail is an important challenge that companies struggle with.

In summary, explainability is a fundamental ethical principle for guiding responsible research in the application and development of AI-systems, is the most pressing ethical challenges for businesses in the ICT-industry, and is according to our analysis as well as to feedback from the industry, insufficiently addressed in the current code of ethics.

6.1.1.3 *Safety*

In section 3.2.2.6, we showed that safety and non-maleficence are not only important for AI ethics but are generally considered important norms in western societies. They are one of the four 'prima facie' duties that were described by W.D. Ross (Ross, 2002). In literature, safety and non-maleficence are often used interchangeably in the context of AI-ethics. However, technically speaking, non-maleficence refers to the obligation to ensure safety, or avoid or minimize harm (Stahl & Wright, 2018). In the context of AI, academics emphasize the need for safety and security and the notion that AI should always avoid foreseeable or unintentional harm (Jobin et al., 2019). The difficulty for non-maleficence with AI arises because full autonomy and control are lost, and a responsibility gap exists. The notion is often discussed in the context of AI for medical applications (Anderson et al., 2006; Keskinbora, 2019). But, other examples of maleficence of AI were also discussed and include a twitter bot that became verbally abusive, an adult content filtering system that failed to block inappropriate content, and deadly incidents with autonomous robots and self-driving cars (Yampolskiy & Spellchecker, 2016). It needs no further explanation that ensuring safe AI-systems will be indispensable to protecting our society on all different levels.

However, businesses in the Dutch ICT-industry indicated that safety with AI-systems is a difficult challenge for them and that they know little about how to deal with safety adequately. During the interviews, two out of six interviewees indicated that safety is one of their most pressing challenges. While in the questionnaire 15% expressed safety as the most pressing challenge in the development and application of AI-systems. Although safety is not judged as a pressing challenge as explainability and fairness, these numbers are still noteworthy because even the consequences of a small number of safety breaches could be catastrophic. Businesses expressed the unexpectedness and unpredictability of negative consequences (e.g., harm) as the key difficulty for dealing with safety challenges. Secondly, AI/ML service-providers find it difficult

to ensure if their products are used in a non-maleficent way because of the open-source ecosystem that it exists in.

Being an important challenge for the reasons found in literature, and a difficult one to deal with for businesses in the ICT-sector, one would expect safety or non-maleficence to be present in the current code of ethics from NLdigital. However, no principle could be found in the analysis of the code of ethics in section 3.3.2 that explicitly mentions safety or non-maleficence. Likewise, during the interviews, interviewees also criticized the absence of a safety and security principle. When asked about the importance of such a principle in a code of ethics for AI in the survey, 82% of respondents agreed or strongly agreed that safety and non-maleficence are important topics that should be included in a code of ethics for AI.

In summary, safety and non-maleficence are fundamental ethical principles for guiding responsible research in the application and development of AI-systems, businesses in the ICT-industry find it difficult to deal with these challenges, and there are currently no principles in the code of ethics that provide guidance to organizations.

6.1.2 Other notable findings

On top of the main findings of this research (i.e., those that answer the main research question), this study has evaluated a wider range of ethical aspects of AI-systems in the context of businesses in the Dutch ICT-sector. In sub-section 6.1.2.1, other notable findings that relate to the importance, adoption, and adherence to the current code of ethics are provided. Finally, in sub-section 6.1.2.2, findings of the ethical governance of AI-systems and concrete measures for dealing with AI-ethics and ethical challenges are concluded.

6.1.2.1 Code of ethics

A large body of the academic literature suggests that if ethical codes are designed and implemented adequately, they may have a significant impact on influencing the ethical behavior of people in an organization (Adam & Rachman-Moore, 2004; Erwin, 2011; Paine, 2004). Likewise, for artificial intelligence, codes of ethics are necessary to guide the ethical behavior of professionals and businesses (Boddington, 2017). Besides positively influencing ethical behavior, codes of ethics can help to overcome the different perceptions and (cultural) norms of people within an organization and help businesses to enhance their public perception and appearance. Companies with a high-quality code of conduct typically rank significantly higher in corporate social responsibility and ethical rankings (Erwin, 2011). Aside from academic literature, businesses also indicate that they believe a code of ethics is necessary to encourage ethical behavior with AI (section 5.5.3). More than half of the large organizations that use AI in the industry already publish their own codes (section 5.4), but the results of the survey suggest that SMEs rarely have their own written down ethical principles. This result strongly emphasizes the importance of 'independent' codes of ethics, guidelines, and principles for such organizations. Evidently, NLdigital's code of ethics can have tremendous value for businesses in positively influencing ethical behavior with AI. The quality, completeness, and connection of the code with the most pressing practical challenges of the industry was the key theme of this work. However, at least some empirical data from the interviews and questionnaire suggests that there are other factors constraining the value that such companies can reap from the code. Most importantly, we found that NLdigital's code, as well as other available codes and guidelines (e.g., European

Union, IEEE), are rarely used by businesses, and especially small and medium-sized businesses. Most businesses, both large and small, were even unfamiliar with NLdigital's code of ethics. The low rate of familiarity suggests that communication has fallen short. This issue will need to be simultaneously addressed to improve the quality of the code of ethics as discussed in the main findings (6.1.1) to achieve more widespread adoption among businesses. Existing scientific literature contributes several requirements for successful implementation of codes of ethics within organizations (Stevens, 2008; Webley & Werner, 2008):

- The code should be embedded in the organizational culture.
- The code should be communicated effectively between the different layers of the organization.

Despite the low adoption or 'familiarity' of the code of ethics, most of the existing principles are carried out to some extent (section 5.5.2). This indicates 1) that the current principles are achievable for many of the businesses in the sector and 2) that companies deem these principles important and are actively working on these 'ethical' activities.

6.1.2.2 *Ethical governance and measures for dealing with ethics*

In all three parts of this research (i.e., literature review, interviews, and questionnaire) the need for concrete measures to encourage, enforce, or assist ethical behavior and dealing with ethical challenges has played a prominent role. In response to sub-research question 5 (i.e., what concrete measures do Dutch ICT-businesses currently use to deal with the ethical aspects of developing and applying AI-systems?) I researched three different classes of measures and mechanisms, i.e., challenge specific measures, internal measures to encourage and facilitate ethical behavior, and pillars of ethical governance. In terms of challenge specific measures, Dutch ICT-businesses mentioned different tools and ways of working for dealing with fairness, explainability, and safety. For creating fair AI-systems, large organizations frequently use software kits to test for fairness in data, specifically for testing biases in sub-populations. SMEs, on the other hand, are mostly reliant on 'manual' checking of the data for intrinsic bias towards sub-populations. Lastly, some businesses indicated to work in diverse teams consisting of people from different backgrounds, ages, gender, and ethnicity. Second, in terms of explainability, businesses can try to build the ability to explain to users what is happening in the software tools or services, e.g., a configurable filter, a list of 'reasons' for why you get a certain recommendation or a complete log of the system. Another way that was found was by actively providing information to their users, for example, by educating clients and explaining the AI-system in a continuous client – company interaction.

Subsequently, section 4.3.2 presented an extensive list of "internal measures' to encourage and facilitate ethical behavior that companies quoted during the interviews. The questionnaire tested which of these mechanisms are predominantly used in the industry. The results indicate that a total of five measures are present within more than half of the businesses. The most important ones being:

1. An open culture to discuss AI and ethics
2. Collaborate actively with the client and/or customer
3. A feedback mechanism for users

4. Strong ethical leadership in the organization
5. Peer review each other's work

These results suggest that companies are in fact actively trying to stimulate ethical behavior within their AI-practices. However, these mechanisms are relatively unmeasurable and 'intangible' mechanisms. More tangible and demonstrable measures are the pillars of ethical governance, which form a framework for solid ethical governance of AI (Winfield & Jirotko, 2018). From the interviews and questionnaire results, it is concluded that it is precisely these measures that small and medium-sized enterprises are often missing. Even though only a small part (20%) of larger organizations possess all five pillars, these companies have vastly more ethical pillars on which they built their ethics practices than SMEs. In total, 67% of all SMEs possess zero pillars and none of the SMEs had more than two out of the five pillars of ethical governance in place.

Considering 1) the large amount of businesses that have started to work with AI-systems, 2) the scalability and relatively large impact of AI-technology, and 3) the absence of concrete pillars, in particular, a code of ethics, a whistleblower mechanism, ethics and RRI training, ethical risk assessments, and transparency about ethical governance, this is cause for concern. Nonetheless, the state of ethical governance within SMEs is not surprising given the early stage these companies are often in and the limited amount of resources they have relative to their larger counterparts. It does not necessarily follow from this conclusion that SMEs should vent all their future efforts on dealing with the ethical aspects of AI, but at the very least asks for awareness of the importance of these aspects. We recognize that not all SMEs will be able to build up these pillars by their own exertion, and thus emphasize the importance of cooperation and joint efforts with other SMEs, organizations such as NLdigital, regulators such as the Dutch government, and large businesses in the industry. Cooperation will be especially appropriate for ethics and RRI training, whistleblower mechanisms, templates for risk assessments, and written down ethics principles. NLdigital's code of ethics remains exemplary of such cooperation.

Sub-research question 7: What concrete measures are missing that would assist small and medium sized companies in dealing with ethical challenges, and how does a code of ethics play a role here?

Answer: This research has focused on two different classes of measures, i.e., measures to encourage and facilitate ethical behavior, and pillars of ethical governance. The most important framework that was found that prescribes measures that should be present within businesses to enforce or govern ethical behavior with AI is that of Winfield & Jirotko (2018). Our research has shown that around half of the large organizations and almost all SMEs are missing important 'mechanisms' that assist or are necessary in dealing with ethical challenges. First, few organizations publish their own code of ethics or ethics principles, or actively use a list of principles created by others. Much academic literature discussed in section 3.2.4 has emphasized the important of such a code or principles. Hence, this leads us to the conclusion that a code of ethics is an important concrete measure that is currently missing. Second, academics have pointed out the necessity and value of whistleblowing channels for AI (section 3.2.4.5). Such channels are currently lacking in the industry; therefore, we conclude that this too in a key measure that is lacking. Following the same line of reasoning, we argue that ethics and RRI training, ethical risk assessments, and the systematic engagement of all relevant stakeholders are important measures that are missing and would assist SMEs in dealing with ethical challenges.

6.2 Limitations of this research

Although this research provides strong evidence for the conclusions presented in section 6.1, there are some important limitations of this study that should be acknowledged. First, there are some of the inherent limitations of the mixed-method research approach that were discussed in section 2.1.1. Because of the limited resources (mainly time) for this study, this study was limited to relatively small sample sizes, which is $n=6$ for the semi-structured interviews and $n=34$ for the online questionnaire. This poses limitations to the generalizability of this research for the entire population. This effect is partly mitigated however by the indication of similar results from both the interviews and the questionnaire, and results that are in line with what we can reasonably expect from literature. Another factor affecting the generalizability of this work is the fact that we only used NLdigital's and the KNVI's members as a sample frame for our analyses. However, this is believed to form a minor challenge in this research as NLdigital's and the KNVI are known to be a representative cross-section of the ICT-industry. Furthermore, discussions with experts in the field also indicated that these are acceptable results.

With regard to the questionnaire specifically, we acknowledge the possibility of inflated results with regards to ethical practices within organizations and AI-adoption in the industry. First, organizations could have indicated to be more ethically driven than they are in reality. To minimize this, minimal retraceable characteristic information was asked of organizations. Only the company name was asked to rule out double responses, and of this, companies were assured that it would be removed in further analyses and publications. Self-inflation of the results is, therefore, expected to be minimal. Furthermore, there is a possibility that the questionnaire was mainly opened by organizations that are working or interested in AI. This was mitigated by making explicit in the invitation of the questionnaire that it is also for companies not working with AI-systems. The invitation featured two large buttons with 'yes' and 'no' for organizations working with AI and not working with AI respectively.

Another limitation of this study, in general, is the restriction to practitioners and experts only from the ICT-industry. We acknowledge that this causes a low external validity of our research, meaning that these results cannot be extrapolated with certainty to other industries. They may still, however, provide an initial framework for ethical challenges and governance of AI in other industries. Further research, however, is needed to determine which industry-specific adjustments are necessary. A second general limitation of this work results from the relatively wide scope that is used in the descriptive assessment of ethical challenges and ethical governance of AI.

We recognize that we have only scratched the surface of concepts such as fairness, explainability, and safety. A narrower scope could bring up more subtleties in the core concepts and identify more subfactors (e.g., security, reliability and non-maleficence for safety). However, for the purpose of this research, this was not necessary because this depth was not required for answering the main research question. Lastly, more detail in the mechanisms to increase ethical behavior could be distinguished, for example, what kind of subject matter is discussed in the ethics training that companies offer. Similarly, more in-depth analyses may be needed into additional mechanisms, governance structures, and the three identified key ethical challenges,

i.e., privacy, explainability, and fairness. For the purpose of this research, however, a wider scope was considered to be sufficient.

6.3 Further discussion

This research has led to some significant expected and unexpected findings. In section 6.3.1 we provide a brief normative ethical reflection on this research and the three ethical aspects that were identified in response to the main research question of this work. Subsequently, in section 6.3.2, we contextualize our other findings with previous research.

6.3.1 Normative ethical reflection on the main findings

This research was focused on establishing the ethical challenges that are currently missing or addressed insufficiently in the code of ethics from NLdigital but are fundamental for guiding responsible research and development in the use and application of AI-systems for the Dutch ICT-industry. To answer this research question, we mapped out the current AI-practices, ethical challenges, ethical governance, and general viewpoints on the code of ethics, making the nature of this work very descriptive. Throughout this work, we recognized different normative viewpoints but have not provided our own position with regards to how the next iteration of a code of ethics should look. However, to provide some direction of where the results of this research could go, this section will briefly discuss some initial normative ethical considerations that have come up during the process. Specifically, we discuss three possible new principles (i.e., for explainability, fairness, and safety) to replace the existing ones. It should be noted that these are suggestions and that the actual definition of these principles will require further research and normative argumentation.

Throughout this work, we have referenced Ross (2002) prima facie duties to show the importance of ethical principles such as justice, non-injury, and harm prevention. We share the view of Anderson and Anderson (2007) that a prima facie form of moral intuitionism is a useful approach for thinking about AI-ethics. The research of Anderson and Anderson is widely recognized as one of the key works in the field of machine ethics. Although their reasoning is primarily focused on creating machines that are explicit ethical agents, we believe the same line of argumentation holds for AI-ethics from a perspective where some form of human intervention is still present, e.g., most of the AI-systems in the Dutch ICT-sector. In short, Anderson and Anderson (2007) and Ross (2002) both share the perspective that “the best approach to ethical theory is one with several prima facie duties some concerned with the consequences of actions and others concerned with justice and rights” (Anderson & Anderson, 2007, p. 18). Such an intuitionistic approach acknowledges the complexity of ethical decision making for AI-systems better than classical deontological theories such as Kant’s duty ethics. In contrary to Kant’s approach where the categorical imperative can justify overlooking the consequences of a decision, the prima facie duties are more ‘guidelines’ that AI-practitioners should aspire to realize but are not treated as all-commanding imperatives. That means, “they could be overridden in on occasion by stronger obligations” (Anderson & Anderson, 2007, p. 18). This approach of not being ‘blind’ to the consequences of ethical decision making with AI is in my opinion vital in the early stages of widespread artificial intelligence where consensus on ethical AI-principles still has to be reached.

During this research, I have also learned the importance of having concrete activities or measures connected to principles, specifically for SMEs. In contrary to large organizations, these

organizations have relatively little knowledge, frameworks, and tools to think about and execute overly abstract principles. We, therefore, envision a combination of principles that are in line with Ross's prima facie duties but also consist of a more practical element. Following this approach, below we provide three examples of possible principles for the code of ethics, where the concrete activities that we recommend are written in normal text, and the ethical principle that should be strived for highlighted in *italic*:

1. **Fairness:** Analyze and monitor your initial data and results, and actively work with a diverse group of people and stakeholders, including people from different backgrounds, gender, and ages *to minimize bias, distribute the benefits and burdens, and treat others impartially equal.*
2. **Safety and/or non-maleficence:** Build a safety net, check the legitimacy of clients, validate your results, and have human checks in the loop *to not harm others physically or psychologically and prevent harm to others from causes other than him- or herself.*
3. **Explainability:** Create extensive logs and documentation, educate stakeholders by teaching or tutorials in different levels of detail, and actively seek to avoid making black-box systems *to provide the opportunity to all relevant stakeholders (e.g., clients and affected individuals) to understand decisions made by an AI-system.*

6.3.2 Discussion on ethical governance of AI within organizations

In section 6.1.2.2 we presented the conclusions to a descriptive analysis of ethical governance within the Dutch ICT-industry. In doing so, we add an applied perspective to Winfield and Jirotko (2018) framework of pillars of ethical governance. In that regard, our work could be treated as a case study of the framework. One of our initial expectations was that only a small amount of businesses in the sector would use NLdigital's or another code of ethics. Furthermore, we expected businesses to just be touching the surface of AI-ethics and its governance. Our research found expected but striking results with regards to the current state of ethical governance within organizations, especially within SMEs. If the framework of five ethical governance pillars is indeed a suitable benchmark for solid ethical governance (section 3.2.3.2), then the majority of the organizations in the industry is lacking fundamental aspects for good ethical governance. Although I share the value of the framework and strongly support the ambition for all businesses working with AI to structure their ethical governance accordingly, one could ask oneself if the pillars are not overly ambitious for the state of technology that we are currently in. Although small and medium-sized enterprises cannot neglect their responsibility to treat AI-ethics with caution, it is understandable that they have less solid ethical governance than large multinationals (e.g., due to their relatively limited amount of resources). Our position is that while society is still working out the preferred ethical pathway for artificial intelligence systems, a maturity or development element should be added into consideration of the evaluation of ethical governance. This does not mean that SMEs can be excused from governing AI ethics, it is merely a recognition that currently not all SMEs will be able to build up these pillars by their own exertion. Cooperation (e.g., training, whistleblowing channels, training) and knowledge transfer (e.g., ethical challenges, tools, and mechanisms) currently look like a logical choice to more solid ethical governance for such organizations. The next section (6.4) will develop some more concrete recommendations of what this cooperative approach could look like.

6.4 Recommendations

Building on the different analyses, findings, and discussions in this dissertation, we propose a few recommendations for the code of ethics and its path forward in section 6.4.1. In section 6.4.2, we propose different additional mechanisms and measures and the way in which they could be attained. Although I believe these recommendations are a step in the right direction, they will require further study.

6.4.1 The code of ethics

Following the main finding of this dissertation, I recommend adjusting the code of ethics based on the three identified areas of improvement. The first principle of the code of ethics should be improved with regards to explainability, the third principle of the code should be improved with regards to fairness, and finally, a safety principle should be added to match the pressing ethical challenges of the industry. Although formulating these new principles requires further extensive normative argumentation, in section 6.3.1, we provided three examples of possible principles for the code of ethics that consisted of both an ethical principle as well as concrete activities that can be done to realize that principle. The notion of specifying more concrete measures and activities that lead to the realization of the principles leads to our second recommendation, i.e., the code should provide more concrete guidance. Although we recognize that a code of ethics is often an abstract value statement, businesses in the industry (mainly SMEs) are in need of more concrete instructions. Our suggestion to do this would be by creating a second more concrete layer under the abstract principles, create an additional guideline, or draft an ethics checklist.

One thing that became clear during this research was that a large part of the interviewees and questionnaire respondents were unfamiliar with NLdigital's code of ethics. The target sample for our study consisted of the employees that actually work with AI-systems on a day to day basis and have to actively deal with some of its inherent ethical challenges. The fact that many of these people have never heard of the code of ethics shows that the code is not penetrating the targeted layers of organizations. On the one hand, this is understandable because NLdigital is an employer organization that mostly communicates with the higher management of businesses. However, the result of this way of communication is that the code of ethics does often not pass further down than the leadership of an organization. Therefore, we recommend communication around the code of ethics to be targeted across all layers of organizations in the industry. Apart from NLdigital, organizations also play a role in effectively diffusing the code throughout the organization. Adoption of the code would be optimal if the organizational culture welcomes change and advocates for ethical behavior. Finally, training and education should be available around the code of ethics either by NLdigital as the creator of the code or by businesses individually. This way companies have the opportunity to ask for specification, learn the more concrete measures that can help in reaching the principles, and become highly familiar with the important ethical aspects of artificial intelligence.

6.4.2 Additional mechanisms

Besides areas of improvement for the code of ethics, throughout this study, we have also identified shortcomings in ethical governance and concrete measures to effectively deal with ethics and ethical challenges. Four concrete recommendations follow our findings (6.1.2.2):

- Establishment of an objective anonymous channel for reporting moral misconduct of artificial intelligence systems

The importance of having channels for reporting misconduct with artificial intelligence has come forward in different publications discussed before (section 3.2.4.5). Such misconduct can take the form of an AI-system acting immorally by itself, or people taking advantage of the technology by allowing moral misconduct to be implemented in AI-systems. Furthermore, whistleblowing is an important method to raise attention to breaches of a code of ethics or other misconduct. The results of our analyses found that only a small number of the industry has an internal whistleblowing channel. Especially SMEs seem to lack the will and resources to establish such internal channels. On the one side, this is threatening for society because moral misconduct could go unreported, on the other side it is also harmful to these very same companies because whistleblowers could go to the media, government, or law enforcement. Therefore, we recommend the establishment of an ‘overarching’ independent whistleblower channel where ICT-organizations can raise concerns about artificial intelligence. Like Bowden (2008), I believe that professional and trade associations have a strong position to facilitate such channels for their respective industries because of the network and credibility they provide to their member companies. Although limited in their capabilities to provide ‘tough’ sanctions when concerns are dismissed, they could still take up penalties to the extent of retraction of a company’s membership. However, professional and trade associations are often partially or fully funded by their members and can thus have a strong aversion of imposing membership withdrawals. Still, we agree with Bowden (2008) that “industry associations and professional societies are the strongest arbiters of behavior in an industry, and it behooves them to apply sanctions that commensurate with this responsibility”. The arguments provided by Bowden (2008) provide an interesting case for NLdigital to pick up such a role for the Dutch ICT-industry.

- Creation of an artificial intelligence ethics and responsible research and innovation (RRI) training program

In a similar way, past research has emphasized the importance of ethics and RRI training for all employees of organizations that are developing, implementing, or applying AI (Winfield & Jirotko, 2018). Our results indicate that very few SMEs provide such training, let alone for everyone in the organization. Therefore, we recommend the creation of an artificial intelligence ethics and RRI training program. One way this can be achieved is through cooperation and joint efforts with other SMEs or larger organizations in the industry. Another possibility is for the program to accompany NLdigital’s code of ethics. In this case, we envision a strong role for NLdigital or branches within the Dutch government here as well.

- Production of templates for ethical risk assessments

Another important part of a pillar for ethical governance that was mostly missing among SMEs was the execution of ethical risk assessments. The current absence of doing such assessments is likely the result of limited resources, know-how, awareness, and availability of easy-to-use tools. To stimulate businesses to undertake ethical risk assessments we suggest that templates for ethical risk assessments be developed for different types of projects or industries. For example, the health sector could develop such templates for AI projects in the health care industry.

- The installation of an objective ethical review board or committee

Although not specifically investigated during this research, we believe that the installation of an objective ethical review board could be beneficial for the industry and society. Businesses, presumably mostly SMEs, could discuss with them their projects, ethical risk assessments, training programs, etc. Such board or committee would then allow businesses to have a second opinion on their work, results, and approach. This board or committee could be seated by for example academic experts or professionals in the field.

6.5 Future research

In section 1.5 the scientific, industrial, societal, and political contributions of this research were laid out. Although AI-ethics is starting to become a well-established field of academic research, there is still much work to do to fully understand all the different aspects of AI-ethics and its position in society, academy, industry, and the political landscape. Building forward on the brief normative reflection (section 6.3.1) in this research specifically, we recommend further research in the normative considerations in the development of the next iteration of the code of ethics. This research has identified fairness, explainability, and safety as the three major areas for improvement and briefly motivated a first thought in the direction principles for these areas should go. However, the formulation of these principles requires deeper ethical reflections and normative argumentation.

A second direction for future research resulting from this work is that of doing a similar descriptive study for different industries. By doing so for the different industries in which AI is used, we can find out which are shared ethical challenges among industries and which challenges differ. This could help determine whether this code of ethics can be extrapolated to other industries and if so where it should be adapted.

Another possibility for future academic research is a replication of this work in a few years' time. It could then be established if and how the predominant challenges in the industry have shifted, if and what additional mechanisms are adopted by multinational companies and small and medium-sized enterprises, and what the impact of different AI-policies has been.

This research has recommended the development and provision of additional mechanisms (e.g., training, whistleblowing structures, and ethical risk assessments), especially for small and medium-sized enterprises in the ICT-sector. More research is unquestionably needed to develop these mechanisms and tools to encourage ethical behavior with AI. These kinds of studies could focus on answering questions such as 1) what should an AI-whistleblowing channel look like and how is it different from traditional structures? 2) What needs to be included in AI-ethics training, to who should it be available, and on what level of understanding does it need to be taught? 3) is it possible to develop a substantial toolkit for anticipating, preventing, and dealing with the ethical challenges of AI, and who should do so? 4) are these means, in fact, effective in influencing ethical behavior with artificial intelligence? 5) how can codes of ethics for artificial intelligence best be implemented and disseminated through all layers of an organization? Lastly, much academic research will be needed in the development of AI-regulation. Researchers and even organizations active in the industry have advocated for good coherent AI-policy and regulation from policymakers. As mentioned before throughout this research, we are also of the opinion that such

regulation will be necessary eventually, but believe it is vital for the technology, all industries, and society that such regulation is right the first time. For this to happen we will among other things need a better understanding of what such regulation could entail and for which ethical norms it may be necessary. NLdigital's code of ethics, and the wide range of other guidelines and codes discussed in this study (section 3.3), can be an important predecessor for such regulation. Researchers will have to study what precisely can be learned from such codes of ethics and how they could be translated into coherent AI-policy.

6.6 Link of this work with the Management of Technology program

The goal of the 'Management of Technology' master program is to educate students as technology analysts in highly technology-based and competitive environments in a variety of industrial sectors. This research fits seamlessly with this goal, as in this research the ICT-sector in the Netherlands is analyzed with regards to ethics and AI. It deals with the socio-cultural system in which a highly specialized new technology such as artificial intelligence is embedded. More specifically, this work has focused largely on how AI-ethics is embedded in the organizational context. This required a thorough understanding of the technical characteristics, possibilities, and limitations of artificial intelligence and machine learning, a comprehensive grasp of ethics and ethical governance, as well as recognition of the social environment and position in the business chain of technological firms. This is exactly what a 'manager of technology' is educated to do, i.e., bring together both technical expertise and a deep understanding of the social and organizational context that such technology lives in. The importance of this program, and the need for more of such 'technology managers' in the context of far-reaching novel technologies follows from the value that such perspective has provided for this dissertation.

References

- ABN-AMRO. (2019). *IT-branche in beeld*. Retrieved from <https://insights.abnamro.nl/2019/03/it-branche-in-beeld/>
- Adam, A. M., & Rachman-Moore, D. (2004). The methods used to implement an ethical code of conduct and employee attitudes. *Journal of Business Ethics*, 54(3), 225-244.
- Adixon, R. (2019). Artificial Intelligence Opportunities & Challenges in Businesses. Retrieved from <https://towardsdatascience.com/artificial-intelligence-opportunities-challenges-in-businesses-ed2e96ae935>
- AINED. (2018). *AI voor Nederland*. Retrieved from https://info.microsoft.com/WE-DIGTRNS-CNTNT-FY19-09Sep-27-ArtificialIntelligenceinNetherlands-MGC0003152_01Registration-ForminBody.html
- Alexander, L., & Moore, M. (2007). Deontological ethics. *The Stanford Encyclopedia of Philosophy*.
- Altman, M., Wood, A., & Vayena, E. (2018). A harm-reduction framework for algorithmic fairness. *IEEE Security & Privacy*, 16(3), 34-45.
- Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. *AI magazine*, 28(4), 15-15.
- Anderson, M., Anderson, S. L., & Armen, C. (2006). *MedEthEx: a prototype medical ethics advisor*. Paper presented at the Proceedings Of The National Conference On Artificial Intelligence.
- Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep machine learning-a new frontier in artificial intelligence research. *IEEE computational intelligence magazine*, 5(4), 13-18.
- Assembly, U. G. (1948). Universal declaration of human rights. *UN General Assembly*, 302(2).
- Baylé, M. (Producer). (2019). Ethical dilemmas of AI: fairness, transparency, collaboration, trust, accountability & morality. *UX Collective*. Retrieved from <https://uxdesign.cc/ethical-dilemmas-of-ai-fairness-transparency-human-machine-collaboration-trust-accountability-1fe9fc0ffff3>
- Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*: Springer.
- Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence*, 316, 334.
- Bowden, P. (2008). Making codes of ethics meaningful and effective. *Journal of Chartered Secretaries*, 60(10), 584-588.
- Bryson, J., & Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, 50(5), 116-119.
- Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3), 294-320.
- Chouldechova, A., & Roth, A. (2018). The frontiers of fairness in machine learning. *arXiv preprint arXiv:1810.08810*.
- Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.): Sage publications.
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *San Fransico, CA: Reuters*. Retrieved on October, 9, 2018.
- Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3-4), 197-387.
- Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology*, 20(1), 1-3. doi:10.1007/s10676-018-9450-z

- Došilović, F. K., Brčić, M., & Hlupić, N. (2018). *Explainable artificial intelligence: A survey*. Paper presented at the 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO).
- Erwin, P. M. (2011). Corporate Codes of Conduct: The Effects of Code Content and Quality on Ethical Performance. *Journal of Business Ethics*, 99(4), 535-548. doi:10.1007/s10551-010-0667-y
- Etaati, L. (2019). Introduction to Machine Learning. In *Machine Learning with Microsoft Technologies: Selecting the Right Architecture and Tools for Your Project* (pp. 3-14). Berkeley, CA: Apress.
- General Data Protection Regulation, L 119/1 C.F.R. (2016).
- Faggella, D. (2019). What is Machine Learning? Retrieved from <https://emerj.com/ai-glossary-terms/what-is-machine-learning/>
- Fairness. (2019). <https://www.lexico.com/en/definition/fairness>: Lexico Publishing Group, LLC.
- Farrell, B. J., & Cobbin, D. M. (1996). A content analysis of codes of ethics in Australian enterprises. *Journal of Managerial Psychology*, 11(1), 37-55.
- Flasiński, M. (2016). History of Artificial Intelligence. In *Introduction to Artificial Intelligence* (pp. 3-13). Cham: Springer International Publishing.
- Floridi, L. (2010). *The Cambridge handbook of information and computer ethics*: Cambridge University Press.
- Gadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. Retrieved from <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing*, 21(6), 58-62.
- Goldsmith, J., & Burton, E. (2017). *Why teaching ethics to AI practitioners is important*. Paper presented at the Workshops at the Thirty-First AAAI Conference on Artificial Intelligence.
- Goode, L. (2018). Google CEO Sundar Pichai compares impact of AI to electricity and fire. Retrieved from <https://www.theverge.com/2018/1/19/16911354/google-ceo-sundar-pichai-ai-artificial-intelligence-fire-electricity-jobs-cancer>
- Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. *AI magazine*, 38(3), 50-57.
- Google. (2018). Machine Learning Fairness. Retrieved from <https://developers.google.com/machine-learning/fairness-overview/>
- Google. (2019). AI at Google: our principles. Retrieved from <https://www.blog.google/technology/ai/ai-principles/>
- HLEG, A. (2019). Ethics guidelines for trustworthy AI. In: Retrieved from High-Level Expert Group on Artificial Intelligence(AI HLEG
- Hunt, T., Song, C., Shokri, R., Shmatikov, V., & Witchel, E. (2018). Chiron: Privacy-preserving machine learning as a service.
- Hussien, M., & Yamanaka, T. (2017). Whistleblowing at work Can ICT encourage whistleblowing? *Joho Chishiki Gakkaishi*, 27(2), 150-154.
- IBM. (2018). *Everyday Ethics for Artificial Intelligence*. Retrieved from <https://www.ibm.com/watson/ai-ethics/>
- IEEE. (2016). Ethically Aligned Design. *IEEE Standards v1*.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1-11.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), 14-26.
- Kagan, S. (2018). *Normative ethics*: Routledge.

- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15-25. doi:<https://doi.org/10.1016/j.bushor.2018.08.004>
- Keskinbora, K. H. (2019). Medical ethics considerations on artificial intelligence. *Journal of Clinical Neuroscience*, 64, 277-282. doi:<https://doi.org/10.1016/j.jocn.2019.03.001>
- Krippendorff, K. (2004). Reliability in content analysis. *Human communication research*, 30(3), 411-433.
- Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016). Accountable algorithms. *U. Pa. L. Rev.*, 165, 633.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
- Lee, Y.-J., & Park, J.-Y. (2018). Identification of future signal based on the quantitative and qualitative text mining: a case study on ethical issues in artificial intelligence. *Quality & Quantity*, 52(2), 653-667. doi:10.1007/s11135-017-0582-8
- Legg, S., & Hutter, M. (2007). A collection of definitions of intelligence. *Frontiers in Artificial Intelligence and applications*, 157, 17.
- Leikas, J., Koivisto, R., & Gotcheva, N. (2019). Ethical Framework for Designing Autonomous Intelligent Systems. *Journal of Open Innovation: Technology, Market, and Complexity*, 5(1), 18.
- Lewis, T. (2014). A Brief History of Artificial Intelligence. Retrieved from <https://www.livescience.com/49007-history-of-artificial-intelligence.html>
- Li, R., Zhao, Z., Zhou, X., Ding, G., Chen, Y., Wang, Z., & Zhang, H. (2017). Intelligent 5G: When cellular networks meet artificial intelligence. *IEEE Wireless communications*, 24(5), 175-183.
- Lufkin, B. (2017). Why the biggest challenge facing AI is an ethical one. Retrieved from <http://www.bbc.com/future/story/20170307-the-ethical-challenge-facing-artificial-intelligence>
- Ma, L., Zhang, Z., & Zhang, N. (2018). Ethical Dilemma of Artificial Intelligence and its Research Progress. *IOP Conference Series: Materials Science and Engineering*, 392(6). doi:10.1088/1757-899x/392/6/062188
- Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46-60. doi:<https://doi.org/10.1016/j.futures.2017.03.006>
- Manyika, J., & Bughin, J. (2018). The promise and challenge of the age of artificial intelligence. Retrieved from <https://www.mckinsey.com/featured-insights/artificial-intelligence/the-promise-and-challenge-of-the-age-of-artificial-intelligence>
- Mason, E. (2006). Value Pluralism; entry in Stanford Encyclopedia of Philosophy (in Zalta NE, ed.). In: Stanford: Centre for the Study of Language and Information, Stanford
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175-183. doi:10.1007/s10676-004-3422-1
- Metz, C. (2019). A Breakthrough for A.I. Technology: Passing an 8th-Grade Science Test. Retrieved from <https://www.nytimes.com/2019/09/04/technology/artificial-intelligence-aristo-passed-test.html>
- Microsoft. (2019). Microsoft AI principles. Retrieved from <https://www.microsoft.com/en-us/ai/our-approach-to-ai>
- Miller, K. W. (2011). Moral Responsibility for Computing Artifacts: "The Rules". *IT Professional*, 13(3), 57-59. doi:10.1109/MITP.2011.46
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). doi:10.1177/2053951716679679
- Montjoye, Y.-A. d., Farzanehfar, A., Hendrickx, J., & Rocher, L. (2017). Solving artificial intelligence's privacy problem. *Field Actions Science Reports. The journal of field actions*(Special Issue 17), 80-83.

- Mouter, N., & Vonk Noordegraaf, D. (2012). *Intercoder reliability for qualitative research: You win some, but do you lose some as well?* Paper presented at the Proceedings of the 12th TRAIL congress, 30-31 oktober 2012, Rotterdam, Nederland.
- Müller, V. C. (2018). *Philosophy and theory of artificial intelligence 2017* (Vol. 44): Springer.
- Nasiripour, S., & Natarajan, S. (2019). Apple Co-Founder Says Goldman's Apple Card Algorithm Discriminates.
- NLdigital. (2019). *Nederland ICT ethical code for artificial intelligence*. Retrieved from <https://www.nederlandict.nl/ethischecodeai/>
- Olhede, S., & Rodrigues, R. (2017). Fairness and transparency in the age of the algorithm. *Significance*, 14(2), 8-9.
- Owotoki, P., & Mayer-Lindenberg, F. (2007). *Transparency of Computational Intelligence Models*, London.
- Paine, L. S. (2004). *Value Shift: Why Companies Must Merge Social and Financial Imperatives to Achieve Superior Performance*: McGraw-Hill, New-York.
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438-450.
- Plummer, L. (2017). This is how Netflix's top-secret recommendation system works. Retrieved from <https://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like>
- Purdy, M., & Daugherty, P. (2016). Why artificial intelligence is the future of growth. *Remarks at AI Now: The Social and Economic Implications of Artificial Intelligence Technologies in the Near Term*, 1-72.
- Rajaraman, V. (2014). JohnMcCarthy—Father of artificial intelligence. *Resonance*, 19(3), 198-207.
- Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping business with artificial intelligence: Closing the gap between ambition and action. *MIT Sloan Management Review*, 59(1).
- Rao, D. A. S., & Verweij, G. (2017). *Sizing the prize: What's the real value of AI for your business and how can you capitalise?* Retrieved from <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>
- Raschka, S. (2015). *Python machine learning*: Packt Publishing Ltd.
- Roeser, S. (2018). *Risk, technology, and moral emotions*: Routledge.
- Ross, W. D. (2002). *The right and the good [original: 1930]*: Oxford University Press.
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach* (3rd ed.): Malaysia; Pearson Education Limited.
- Salvino, M. (2018). CEOs Must Navigate The Ethics Of AI And Machine Learning. Retrieved from <https://chiefexecutive.net/ceos-navigate-ethics-ai-machine-learning/>
- Sanders, A. K., Falcão, T., Haider, A., Jambeck, J., LaPointe, C., Vickers, C., & Ziebarth, N. (2018). *World Economic and Social Survey 2018: Frontier Technologies for Sustainable Development* (978-92-1047-224-1). Retrieved from New York:
- Santoni de Sio, F., & van den hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, 5, 15. doi:10.3389/frobt.2018.00015
- SAS. (2019). Machine Learning: What it is and why it matters. Retrieved from https://www.sas.com/en_us/insights/analytics/machine-learning.html
- Schwartz, M. S. (2004). Effective Corporate Codes of Ethics: Perceptions of Code Users. *Journal of Business Ethics*, 55(4), 321-341. doi:10.1007/s10551-004-2169-2
- Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach*: John Wiley & Sons.

- Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J. C., . . . Bauer, Z. (2018). *The AI Index 2018 Annual Report*". Retrieved from <https://aiindex.org/>
- Singh, N. (2016). How to get started as a developer in AI. Retrieved from <https://software.intel.com/en-us/articles/how-to-get-started-as-a-developer-in-ai>
- Singh, S., Okun, A., & Jackson, A. (2017). Artificial intelligence: Learning to play Go from scratch. *nature*, 550(7676), 336.
- Sinnott-Armstrong, W. (2019). Consequentialism. *The Stanford Encyclopedia of Philosophy*.
- Smith, B., & Browne, C. A. (2019). *Tools and Weapons: The Promise and the Peril of the Digital Age*: Penguin Press.
- Soares, E., Ahmad, T., Levush, R., Guerra, G., & Martin, J. (2019). Regulation of Artificial Intelligence: The Americas and the Caribbean. Retrieved from <https://www.loc.gov/law/help/artificial-intelligence/americas.php>
- Stahl, B. C., & Wright, D. (2018). Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation. *IEEE Security & Privacy*, 16(3), 26-33.
- Stevens, B. (1994). An analysis of corporate ethical code studies: "Where do we go from here?". *Journal of Business Ethics*, 13(1), 63-69. doi:10.1007/bf00877156
- Stevens, B. (2008). Corporate Ethical Codes: Effective Instruments For Influencing Behavior. *Journal of Business Ethics*, 78(4), 601-609. doi:10.1007/s10551-007-9370-z
- TNO. (2019). Hoe kunnen we artificial intelligence gaan vertrouwen? Retrieved from <https://www.tno.nl/nl/tno-insights/artikelen/hoe-kunnen-we-artificial-intelligence-gaan-vertrouwen/>
- Van de Poel, I., & Royakkers, L. (2011). *Ethics, technology, and engineering: An introduction*: John Wiley & Sons.
- van der Beek, P. (2019). Ict-branche publiceert code voor ai en ethiek. Retrieved from <https://www.computable.nl/artikel/nieuws/digital-innovation/6630597/250449/ict-branche-publiceert-code-voor-ai-en-ethiek.html>
- Venkatesh, V., Brown, S. A., & Sullivan, Y. W. (2016). Guidelines for conducting mixed-methods research: An extension and illustration. *Journal of the Association for Information Systems*, 17(7), 2.
- Webley, S., & Werner, A. (2008). Corporate codes of ethics: Necessary but not sufficient. *Business Ethics: A European Review*, 17(4), 405-415.
- Wilson, M. (2019). This free AI reads X-rays as well as doctors. Retrieved from <https://www.fastcompany.com/90326445/this-free-ai-reads-x-rays-as-well-as-doctors>
- Winfield, A. F., & Jirotko, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180085.
- Wisdom, J., & Creswell, J. W. (2013). Mixed methods: integrating quantitative and qualitative data collection and analysis while studying patient-centered medical home models. *Rockville: Agency for Healthcare Research and Quality*.
- Yampolskiy, R. V., & Spellchecker, M. (2016). Artificial intelligence safety and cybersecurity: A timeline of AI failures. *arXiv preprint arXiv:1610.07997*.
- Yao, M., Zhou, A., & Jia, M. (2018). *Applied Artificial Intelligence: A Handbook for Business Leaders*: Topbots Inc.
- Zheng, Y., Dave, T., Mishra, N., & Kumar, H. (2018). *Fairness in reciprocal recommendations: A speed-dating study*. Paper presented at the Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization.

A. Interview protocol

Introduction

Once again, thank you for your time and availability for this interview. Your input will be very valuable to my research results. My name is Arnoud, and I am a master's student Management of Technology at the TU Delft. The interview we will be conducting today will be used for my master thesis research that I am doing for NLdigital. As you might know, NLdigital released a code of ethics for its members to guide ethical AI practices. After the introduction of the code, they realized that there is little know about AI/Ethics practices and the use of the code in the industry. Based on this new information that will be gathered in this descriptive study, they want to develop the next generation of this code. The purpose of this interview is to learn what your practice looks like, what kind of challenges you encounter, how you deal with those, and what your opinion is on the code of ethics.

Permission

Before we start the interview, I would ask for your permission to record the audio of this interview. Representing your responses correctly will be crucial for the quality of my research. Moreover, this will allow me to have an engaging conversation with you during the interview. Of course, the recording will remain confidential and all the data that is used in the final report will be anonymized, that is neither your name nor your company's name will be referenced. Please take all the time you need to read the 'informed consent form' and sign the form. If you have no objections, I would like to proceed with the interview questions.

Introductory questions

- **IQ1.** How many people does your company employ globally and in the Netherlands?
- **IQ2.** Could you briefly introduce to me what your company does?
- **IQ3.** Could you tell me about the function you have in the company and your daily activities?
 - **IQ3.1.** How long have you had this function?
 - **IQ3.2.** What have you done previously, and how did you roll into this function?

Open questions

- **OQ1.** How does your company implement artificial intelligence (MLaaS or self-developed)?
- **OQ2.** What do you think of AI and Ethics?

Main questions

Q1. What is the most pressing ethical challenge you encounter when designing or implementing AI?

- **Q1.1.** Can you give an example of a situation in which you encountered this challenge?

- **Q1.2.** What are other ethical challenges that you frequently encounter or worry about?
- **Q1.3.** Can you rank these challenges based on impact on your work?
 - Privacy
 - Fairness
 - Maleficence & safety
 - Responsibility
 - Transparency

Q2. What do you, personally, believe to be good practice with regards to approaching and solving the following ethical challenges?

- **Q2.1. Privacy:** What do you, personally, believe to be good practice for approaching and solving privacy challenges that you encounter in the use of AI?
- **Q2.2. Transparency:** What do you, personally, believe to be good practice for approaching and solving transparency challenges that you encounter in the use of AI?
- **Q2.3. Responsibility:** What do you, personally, believe to be good practice for approaching and solving responsibility challenges that you encounter in the use of AI?
- **Q2.4. Maleficence:** What do you, personally, believe to be good practice for approaching and solving safety/maleficence challenges that you encounter in the use of AI?
- **Q2.5. Fairness:** What do you, personally, believe to be good practice for approaching and solving fairness challenges that you encounter in the use of AI?

Q3. How do your colleagues treat these challenges differently?

- **Q3.1** What do you think about your colleague's solution?
- **Q3.2** Would you adopt their viewpoint?

Q4. How do you discuss ethics (e.g. responsibility) with partners or customers in the chain?

Q5. What mechanisms for dealing with ethical challenges does your company have in place?

- **Q5.1.** What other mechanisms would be of value to you and your company?
- **Q5.2.** Does your organization have a whistleblowing mechanism in place (specifically for AI)?

Q6. Do you have a code of ethics or set of principles that you follow?

- **Q6.1.** If NLdigitals Code of Ethics, why do you use this code of ethics?
- **Q6.2.** If they don't use the code; why don't you use the code of ethics?

Q7. What pillars for ethical governance of AI does your company possess (based on Winfield and Jirotko (2018))?

- **Q7.1.** Does your company publish a code of ethics, alongside a 'whistleblower' mechanism? (Q5, Q6)
- **Q7.2.** Does your company provide ethics and responsible innovation training for everyone in the organization?

- **Q7.3.** Does your company actively involve a wide range of stakeholders and carry out ethical risk assessments of new products? (Q4)
- **Q7.4.** Is your company transparent about ethical governance, if so, how? (not just follow the law)
- **Q7.5.** Do you feel every part of your organization takes ethical governance seriously?

Q8. I would like to ask you to take a few minutes to look at the code of ethics.

- **Q8.1.** What are useful principles in the code of ethics for you?
- **Q8.2.** What principles do you not find useful for dealing with ethical challenges?
- **Q8.3.** Are there any important principles missing that would help you?
- **Q8.4.** What do you think of the way the principles are formulated? Are they clear?
- **Q8.5.** How can the ethics code have more impact for you and your company?

Closing remarks

This was my last question, and thereby the end of my interview, thank you very much for your time. Are there any other points or questions that you think should be addressed to capture the AI ethics landscape in the industry and or to improve the code of ethics? Finally, if you are interested, I would be glad to send you a summary of the results and the final thesis of this study once the project is finished.

B. Interview analysis Atlas.ti

Code Report – Grouped by: Code Groups
All (187) codes

AI practice

16 Codes:

- Deliver services
 - Information platform
 - Text analysis
 - People that know how to do ML/stat well are limited
 - Majority ML ecosystem is open-source
 - Machine learning platform
 - Lower entry level of machine learning
 - Text recognition and analysis
 - Recognize and model content
 - Develop own models
 - AI is less advanced than perceived
 - Automating analysis task: due diligence
 - AI means machine learning
 - Customer has issue with data
 - Four step model
 - End-to-end machine learning solutions
-

Code of ethics

10 Codes:

- Personal informal ethics principles
- Code is important for consciousness
- Accessibility of code influences use
- Code has influenced behavior

- Relevancy of principles is dependent
 - Possibility to opt out
 - Own ethics principles
 - Verbal informal code of ethics
 - Misaligned ethics thinking
 - Trained on principles
-

Explainability

8 Codes:

- Difficult to let consumers spend too much time to explain
 - Cultural differences influence explainability
 - Explainability dependent on project
 - Range of considerations is too high to explain well
 - Try to show people what they're doing
 - Customers demand transparency and explainability
 - Explainability as transparency
 - Education for explainability
-

Fairness

5 Codes:

- Explaining fairness is hard, depends on context
 - Fairness and biases should be important for everyone
 - Fairness is dependent on the situation
 - Aware of biases, check from that viewpoint
 - Provide evidence of unbiased system
-

Feedback code NLdigital

32 Codes:

- Provide evidence of unbiased system
- Code is without obligations
- Safety and security missing in code
- Education principle is recognizable
- Analyzing data to prevent bias missing
- Being aware and transparent does not minimize bias
- Missing preventive privacy principles
- Level of abstraction is good
- Feedback principle is important
- Code useful for SMEs
- Include using diverse group of developers and data
- Internal feedback loop for design is missing
- Abstract principles offer little guidance
- Anonyms channel for misconduct
- 1st and third principle are fuzzy
- Definition of equity and inclusivity missing
- 8th principle is good but very hard
- Transparency is strong in code
- Be aware of limitations of using plug-ins
- Guidance on fairness are missing
- Customer dealing with AI is a good principle
- Implementation of method for feedback is difficult
- SMEs use common sense instead of principles
- Relevance of code is case dependent
- Possibility to trace recommendations (5) is very important
- Principles intuitively make sense
- Level of detail of transparency is unclear
- Formulation of first principle is unclear

- Code stuck at leadership
 - Use code if they provide extra value over own principles
 - Test and certify companies on code adherence
 - Build safety net
-

Importance of ethics

6 Codes:

- Many small decisions make a big one
 - Customers interested in ethics
 - Ethics is about impact on society
 - Ethics is key dimension
 - Model unusefull with illegal data
 - First technology, then ethics
-

Increase impact CoE in industry

12 Codes:

- Code stuck at leadership
 - Use code if they provide extra value over own principles
 - Test and certify companies on code adherence
 - Unforeseen impact is important in AI ethics
 - Communication and education
 - Sharing of code throughout organization
 - More concrete practical questions
 - Revoke mark when non-compliant or preventive test
 - Provide training on code
 - Focus on SMEs
 - Too few people in organization for training
 - Code should become a quality mark
-

Interpretation of colleagues

4 Codes:

- Trained on principles
- Ethical judgement is standard practice
- Individual employees values are leading
- Interpretation not streamlined due to different backgrounds

Mechanisms to encourage ethical behavior

19 Codes:

- Feedback loop with ethics checks
- Central storage for all ML models/data
- Fear of rebuttal
- Media scrutiny enforces ethical behavior
- Raise awareness level by publications, documentaries, etc.
- Open culture of discussing
- Involve wide range of stakeholders
- No training, but talks exist
- Open culture about AI ethics
- Discuss ethics with customer
- Feedback mechanism for users
- Collaboration with client
- Seek advice at colleagues and communities
- Punished or fired for unethical behavior
- Ethical leadership
- Peer reviewed work
- Decline project with illegal data
- Check legitimacy of the client
- Whistleblowing channel for AI

Most pressing ethical challenges

4 Codes:

- Explainability
- Fairness - bias

- Privacy
- Safety

Other ethics challenges

5 Codes:

- Business people know the data best
- Operate in niche, different challenges
- Trade off human intervention vs AI
- Demanding ethics from customer may cause the loss of client
- Test plugins (bias, safety, privacy)

Pillars for ethical governance of AI

6 Codes:

- IV3 - Pillars present: None
- IV6 - Pillars present: Wide range of stakeholders
- IV1 - Pillars present: Ethics principles and whistleblowing channel, training for everyone, wide range of stakeholders, transparent ethical governance and ethics taken seriously across organization
- IV 4 - Pillars present: Wide range of stakeholders and risk assessments, everyone takes it serious
- IV2 - Pillars present: Ethics code
- IV5 - Pillars present: None

Pressing ethical challenges

10 Codes:

- Making unbiased models
- Explainability is key challenge
- Freedom of speech is key challenge
- Biases
- Cannot see what customers analyze
- Fairness is key challenge
- Safety is key challenge

- Data collection
 - Wrong predictions
 - Safety for society is important challenge
-

Privacy

7 Codes:

- Data collection
 - Privacy limited to GDPR compliance
 - Cannot oppose extra data governance rules on clients
 - Training model with confidential/unknown data
 - Clients demand privacy
 - GDPR is sword with two sides
 - Collect data that you need
-

Responsibility

6 Codes:

- Acknowledge responsibility is open point
 - Responsibility not perceived as that important
 - Responsibility not yet an issue
 - Responsibility dependent on project
 - Responsibility gap in practice
 - Responsibility is for client – Tool
-

Safety

6 Codes:

- Wrong predictions
 - Safety for society is important challenge
 - Open-source ecosystem limits safety
 - Safety has a non-algorithmic dimension
 - Unexpected outcomes may cause maleficence
 - Transparent model increases safety
-

Transparency

10 Codes:

- Customers demand transparency and explainability
 - Explainability as transparency
 - Struggling with level of understanding of user
 - Clients demand transparency
 - Trade-off transparency and privacy
 - Training to be transparent
 - trade-off how much information you show
 - Too transparent can spill confidential data or reverse engineer
 - Complexity of actors influences transparency
 - AI is a black-box, no transparency planned
-

Ways to deal with fairness challenges

3 Codes:

- Sub population analysis tool
 - Diverse development team for fair systems
 - Fairness kit software
-

Ways to deal with privacy challenges

13 Codes:

- Decline project with illegal data
- Encrypt data of users
- Deal: Discussion with client
- Help customer set up data ops
- Always know how data was acquired
- Deal: Regulation (GDPR)
- Anonymize data
- Data processing agreement
- Delete unnecessary data
- Deal: Training on data/privacy
- Privacy check

- Market pressure and fines
 - Data stewards
-

Ways to deal with responsibility challenges

3 Codes:

- Responsibility is for client - Tool
 - Vocalize responsible AI
 - Creation of clear community standard for content
-

Ways to deal with safety challenges

10 Codes:

- Build safety net
- Check legitimacy of the client
- Transparent model increases safety
- Validate recommendations of system
- Selective data sources; primary source
- Humans as safety check
- Smart algorithms to prevent harm
- Use AI as tool to increase safety

- Validate results to ensure safety
 - Validation by random checks
-

Ways to deal with transparency challenges

8 Codes:

- Education for explainability
 - AI is a black-box, no transparency planned
 - Documentation, logs and authorization in system
 - Simple explainability/transparency functionality in product
 - Teaching and tutorials to explain
 - Explainability built into software
 - Configurable filter for customer
 - Explain the model to be transparent
-

Whistleblowing

2 Codes:

- Whistleblowing channel for AI
- No whistleblower channel, but something to think about

C. Questionnaire

General questions

1. What is the name of your organization? (will be made anonymous)

..... <fill – in>

2. Your organization is headquartered in:

..... <drop-down menu>

3. Footprint of your organization:

- Netherlands
- Benelux
- Europe
- Worldwide
- <fill-in>

4. Number of employees in your organization:

- 0 – 10 employees
- 11 – 50 employees
- 51 – 250 employees
- 250 – 1000 employees
- 1001 – 10,000 employees
- > 10,000 employees

Mapping the AI-landscape

5. Does your organization use artificial intelligence (AI) for internal processes, products, or services?

- Yes
- No

IF NO:

5.1 Is your organization planning to use artificial intelligence (AI) for your processes, products, or services?

- Yes
- No

5.2 What is the reason you are not (yet) using or developing artificial intelligence? (multiple answers possible)

- It is not useful or provides no additional value for our processes, products, or services
- We don't have the resources to do so
- Ethical challenges and potential negative consequences are holding us back

- We cannot attract and don't have the right talent to do so
- < Fill – in >

Proceed to the end of the interview

IF YES:

6. What stage is your company currently in with artificial intelligence (AI)?

- We are currently developing AI-systems, but haven't implemented it yet
- We are currently using AI-systems for some processes, products, or services
- We are in the stage where AI-systems lie at the core of our processes, products or services

7. What do you use artificial intelligence (AI) for? (multiple answers possible)

- We provide AI-enabled services
- We create AI-enabled products
- We use it for internal processes such as marketing and sales, supply-chain management, manufacturing, human resources, corporate finance, etc.
- < fill – in >

8. How does your organization develop AI-enabled processes, products or services?

- We use our own infrastructure and open-source libraries to develop our own AI-enabled products or services
- We don't program or develop any of the AI-systems ourselves but use services of others to develop our systems such as Machine Learning as a Service (MLaaS)
- We develop our own AI-systems, but also use services of others such as AWS, Microsoft Azure, Google cloud, IBM Watson, etc.
- <fill in >

9. What market do you use your AI-enabled services or products for: (multiple answers possible)

- Agriculture
- Automotive
- Energy & environment
- Financial services
- Healthcare systems and services
- High tech
- Manufacturing
- Power and energy sector
- Professional services
- Information and communications technology
- Travel, transport, logistics
- Retail
- Real estate
- Pharmaceutical and medical products
- < fill – in >

Ethics challenges

10. What is/are the most pressing ethical challenge(s) that you encounter in the design or implementation of your AI-enabled products or services? (two answers possible)

- Fairness: prevention of biases that cause 'unfair' outcomes
- Explainability: explaining predictions, recommendations, or decisions
- Safety and non-maleficence: preventing people from harm
- Privacy: collecting and protecting people's data and privacy in a responsible way
- Responsibility: who is responsible and accountable for the consequences of AI-systems
- Transparency: acts of communication about the AI-system
- None: we never encounter ethical challenges
- < fill – in >

11. How important are the following ethical challenges for you in the design or implementation of AI-enabled products or services in your organization?

Table B.1: Likert scale answers for question number 9

	Not important	Slightly important	Moderately important	Important	Very important
Fairness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Explainability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Safety and non-maleficence	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Privacy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Responsibility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Transparency	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

AI – Ethics mechanisms

12. What internal mechanisms for increasing ethical behavior with AI does your organization have in place? (multiple answers possible)

- A feedback mechanism for users
- AI and ethics training
- An open culture to discuss AI and ethics
- Strong ethical leadership in the organization
- We publish publications, videos, and documentaries
- We peer review each other's work
- People that behave unethical will get punished or fired
- We collaborate actively with the client and/or customer
- Communities and colleagues at which we can seek help with ethical challenges
- Central storage system in which we store all AI models and data
- We use a feedback loop with ethics checks throughout the life cycle of projects

- A secure channel where employees can expound their moral concerns about the company's practices
- A written down set of rules or principles by which we all comply
- We don't encourage ethical behavior
- < fill – in>
- < fill – in>
- < fill – in>

13. Which of the ethical governance pillars below does your organization have in place? Tick the boxes that apply to your organization. (multiple answers possible)

- We publish a code of ethics or ethics principles for artificial intelligence
- We have a whistleblowing mechanism where we can raise ethical concerns without fear of retribution
- We provide ethics and responsible innovation training for everyone in the organization
- We engage a wide range of stakeholders within a framework of anticipatory governance
- We undertake ethical risk assessments of all new products, and act upon the findings of those assessments
- We are transparent about our ethical governance process, for example, show on our website memberships of ethics boards, publish ethics case studies, or publish a code of ethics.
- I feel that everyone, both our 'lower-rank' employees as well as top-level management, really values ethical governance of AI in our organization.
- None of the above

Code of Ethics

14. Are you familiar with NLdigital's code of ethics for artificial intelligence?

- Yes
- No

15. How often do you use NLdigital's code of ethics for artificial intelligence?

- Very frequently
- Frequently
- Occasionally
- Rarely
- Never
- I didn't know NLdigital has a code of ethics

16. To what extent do you currently carry out the following activities?

	Not at all	Small extent	Some extent	Moderate extent	Large extent
Make it clear to customers, end-users, and suppliers when they are dealing with AI	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Make clear the responsibility that customers, end-users, and suppliers bear when creating or applying AI	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Be transparent and communicate clearly about the state of AI, its technical possibilities, and limitations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provide insight into the data that is used by the AI application or offer the possibility for customers, end-users and partners to acquire such insight	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provide the possibility for all other parties involved to trace recommendations or decisions made by the AI system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Monitor the behavior of the AI-application and give users of the AI application a means for providing feedback	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Share knowledge and provide education about AI in general within and outside the sector	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provide information to customers, end-users, and suppliers regarding the system and its learning curve and the technology	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provide information to customers, end-users, and suppliers regarding data usage outside the EU	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

17. What is your opinion on the following statements:

Strongly disagree	Disagree	Neutral	Agree	Strongly agree
-------------------	----------	---------	-------	----------------

A code of ethics is necessary to encourage ethical behavior with artificial intelligence	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fairness and the minimization of biases is an important topic that should be included in a code of ethics for AI	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Explainability is an important topic that should be included in a code of ethics for AI	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Safety and non-maleficence are important topics that should be included in a code of ethics for AI	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
More ethics mechanisms (trainings, risk assessments, whistleblowing channels, etc.) should be made available to my organization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

18. Would you adopt NLdigital’s ethics code if it incorporated principles on more practical challenges of small and medium sized enterprises in the next version?

- Yes
- No
- < fill – in >

You have completed the survey. Thank you very much for your time and effort to answer these questions. Your answers are very valuable for the quality of this research. The results will be used to draw a landscape of the state of AI in the ICT-sector and conduct an in-depth analysis of ethical challenges with AI in the industry. For any further questions, or for comments about the content of the survey please contact arnoudnederpel@nldigital.nl, or you could leave your feedback in the box below

..... < fill – in >

To receive the results of the survey, please provide your email-adress below:

..... < fill – in >

D. Data treatment

After all the responses were collected from the online questionnaire, the data was exported from 'Spotler', the online survey environment. Data was exported in CSV format to be cleaned and treated using the Power Query editor of Microsoft Excel. This section explains the different treatment processes that were executed on the data in order to make them workable for further analysis. First, the original variables in the CSV file are renamed in order to increase readability and ease further analysis. Second, unnecessary and unwanted variables are deleted, and an extra indexing variable is created. Finally, test responses, incomplete responses, and outliers are deleted.

D.1. Renaming variables

The CSV data file from 'Spotler' exports the full question-text as written down in the electronic survey as the variables. To create a clearer structure and ease further analysis, these original variables are renamed into new labels and displayed in Table 18.

Table 18: Renamed variables and question number

Original name	New label	Question number
Datum	Date	
Testreactie	TestResponse	
Volledig ingevuld	CompleteResponse	
What is the name of your organization? (will be anonymized)	OrganizationName	
Your organization is headquartered in:	Headquarter	
What is the footprint of your organization?	Footprint	
How many employees do you have in your organization?	Employees	
Does your organization use some form of artificial intelligence (AI) for your products, services, or internal processes?	AI_Use	
Is your organization planning to use artificial intelligence (AI) for your processes, products, or services?	AI_Planning	
What is the reason you are not (yet) using or developing artificial intelligence? (multiple answers possible)	Reason_No_Use	
Please provide your email adress if you would like to receive the results of this research (not used in report):	Email_No_Use	
What stage is your company currently in with artificial intelligence (AI)?	AI_Stage	
What do you use artificial intelligence (AI) for? (multiple answers possible)	AI_Purpose	
How does your organization develop AI-enabled processes, products or services?	AI_Development	
What market do you use your AI-enabled services or products for: (multiple answers possible)	TargetMarket	

What is/are the most pressing ethical challenge(s) that you encounter in the design or implementation of your AI-enabled products or services? (two answers possible)	Pressing_Challenge	
How important are the following ethical challenges for you in the design or implementation of AI-enabled products or services in your organization? - Fairness	Importance_Fairness	
How important are the following ethical challenges for you in the design or implementation of AI-enabled products or services in your organization? - Explainability	Importance_Explainability	
How important are the following ethical challenges for you in the design or implementation of AI-enabled products or services in your organization? - Safety and non-maleficence	Importance_Safety	
How important are the following ethical challenges for you in the design or implementation of AI-enabled products or services in your organization? - Privacy	Importance_Privacy	
How important are the following ethical challenges for you in the design or implementation of AI-enabled products or services in your organization? - Responsibility and accountability	Importance_Responsibility	
How important are the following ethical challenges for you in the design or implementation of AI-enabled products or services in your organization? - Transparency	Importance_Transparency	
What internal mechanisms for increasing ethical behavior with AI does your organization have in place? (multiple answers possible)	InternalMechanisms	
13. Which of the ethical governance pillars below does your organization have in place? Tick the boxes that apply to your organization. (multiple answers possible)	GovernancePillars	
Are you familiar with NLdigital's code of ethics for artificial intelligence?	FamiliarityCoE	
How often do you use NLdigital's code of ethics for artificial intelligence?	FrequencyCoE	
To what extent do you currently carry out the following activities? - Make it clear to customers, end-users, and suppliers when they are dealing with AI	Extent_Principle2_Part1	
To what extent do you currently carry out the following activities? - Make clear the responsibility that customers, end-users, and suppliers bear when creating or applying AI	Extent_Principle2_Part2	
To what extent do you currently carry out the following activities? - Be transparent and communicate clearly about the state of AI, its technical possibilities, and limitations	Extent_Principle3	
To what extent do you currently carry out the following activities? - Provide insight into the data that is used by the AI application or offer the possibility for customers, end-users and partners to acquire such insight	Extent_Principle4	
To what extent do you currently carry out the following activities? - Provide the possibility for all other parties	Extent_Principle5	

involved to trace recommendations or decisions made by the AI system		
To what extent do you currently carry out the following activities? - Monitor the behavior of the AI- application and give users of the AI application a means for providing feedback	Extent_Principle6	
To what extent do you currently carry out the following activities? - Share knowledge and provide education about AI in general within and outside the sector	Extent_Principle7	
To what extent do you currently carry out the following activities? - Provide information to customers, end-users, and suppliers regarding the system and its learning curve and the technology	Extent_Principle8_Part1 23	
To what extent do you currently carry out the following activities? - Provide information to customers, end-users, and suppliers regarding data usage outside the EU	Extent_Principle8_Part4	
What is your opinion on the following statements? - A code of ethics is necessary to encourage ethical behavior with artificial intelligence	Opinion_NecessityCoE	
What is your opinion on the following statements? - Fairness and the minimization of biases is an important topic that should be included in a code of ethics for AI	Opinion_NecessityFairnessInCoE	
What is your opinion on the following statements? - Explainability is an important topic that should be included in a code of ethics for AI	Opinion_NecessityExplanationInCoE	
What is your opinion on the following statements? - Safety and non-maleficence are important topics that should be included in a code of ethics for AI	Opinion_NecessitySafetyInCoE	
What is your opinion on the following statements? - More ethics mechanisms (trainings, risk assessments, whistleblowing channels, etc.) should be made available to my organization	Opinion_NecessityMoreMechanisms	
Would you adopt NLdigital's ethics code if it incorporated principles on more practical challenges of small and medium sized enterprises in the next version?	Willingness_AdoptionCoE_SME	
Please provide your email adress if you would like to receive the results of this research (not used in report):_1	Email_Use	

D.2. Removing and adding variables

A condition that was promised to the respondents so that they would fill out the questionnaire truthfully without fear of giving away confidential information or other negative consequences, was to anonymize the data for analysis and publication. Variables that would lead to direct identification of organizations or individuals are therefore removed from the dataset. Table 19 displays the removed variables. One index variable is added to ease referencing during the analysis process (Table 20).

Table 19: Removed variables

Removed variables	Remarks
OrganizationName	Data anonymization
Email_No_Use	Data anonymization
Email_Use	Data anonymization

Table 20: Added variables

Added variables	Remarks
Index	Added an index column starting from 1
Size	Distinction between SMEs (<=250 employees) and large organizations (>250 employees)

D.3. Cleaning of test responses, incomplete responses, and outliers

The last step of the treatment process functions to remove any cases that could pollute the data or analysis. First, test cases were removed immediately in the online software tool. All values for TestResponse remaining in the document were 'N', indicating that indeed no test response values were remaining in the final data set. Then, the dataset was analyzed for any incomplete responses. A total of 12 incomplete responses were found. These responses recorded complete responses up until question 5 (i.e., does your organization use artificial intelligence (AI) for internal processes, products, or services?). These 12 responses were deleted from the dataset.

The total amount of valid respondents remaining after question 5: $46 - 12 = 34$.

Lastly, the dataset was analyzed for any outliers, irrational responses, and anomalies. As expected, (because the respondents are all community members of NLdigital) no such responses were recorded, and no additional responses are deleted from the dataset. In the end, the remaining dataset included 34 valid responses and 40 variables.