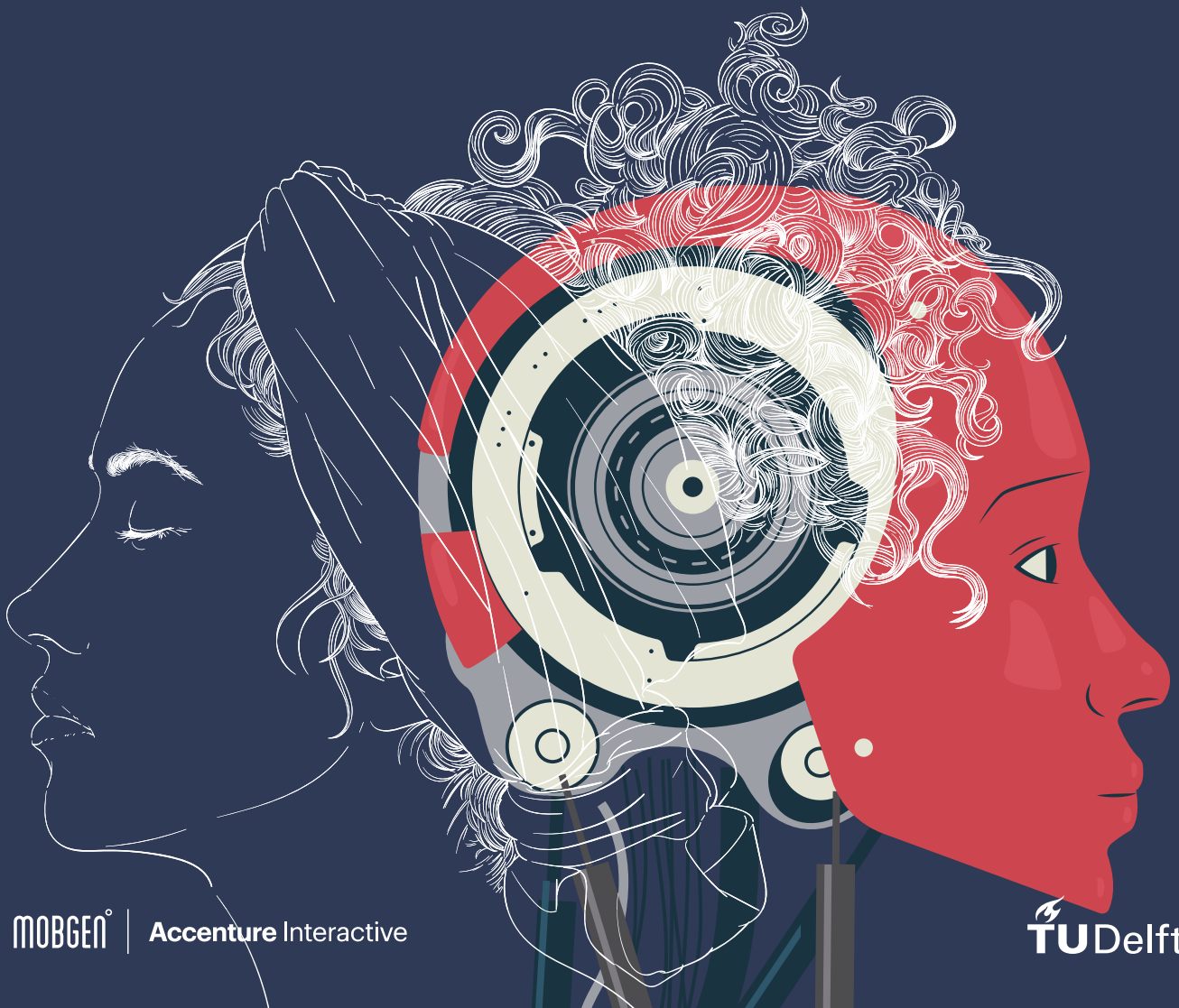
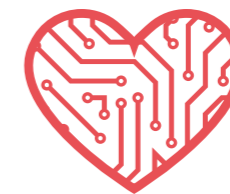


Design of an Ethical Toolkit for the Development of AI Applications

Master Thesis
Strategic Product Design
Mario Alberto Sosa Hidalgo



Design of an Ethical Toolkit for the Development of AI Applications



Author

Mario Alberto Sosa Hidalgo
mariososadi@gmail.com

Master thesis

MSc. Strategic Product Design
Faculty of Industrial Design Engineering
Delft University of Technology

Supervisors

Chair - Dr. Rebecca Price
Faculty of Industrial Design – Product Innovation Management

Mentor - M.Sc. Nick Sturkenboom
Faculty of Industrial Design – Product Innovation Management

Company mentor - Sebastian Veldman
Innovation Lead at MOBGEN | Accenture Interactive

August, 2019

** The cover was designed using illustrations from Freepik.com

Master Thesis
Mario Alberto Sosa Hidalgo

Acknowledgments

The report that you are reading right now represents one of the biggest achievements of my life. Not only because of the academic challenge that it represented, but also for the emotional and social events that I experienced during this stage of the masters. Throughout this journey, I had the opportunity to meet incredible people that supported me immensely. I would like to thank every single one of them because, without them, this project would have been only a sentence in a lost notebook.

I would like to thank first of all my supervisory team. Rebecca, thank you for your time and enthusiasm about my project, especially for those times where you convinced me to relax, take a break, and then come back strong. This report is proof that enjoying the process is part of a graduation project. Nick, thank you for your exceptional coaching, for always suggesting different approaches to my problems, and for your valuable time during our meetings. Without your advice and support, this project would not have been the same. I think I am ready for a P.h.D, now that I know what it is all about. Sebastian, thank you for all the support within MOBGEN, your help with Fjord, and also for letting me paving my path around this complicated but exciting topic. I enjoyed working with all three of you as you always make me feel confident about myself and motivated to keep going.

Furthermore, I enjoyed working at the LAB in MOBGEN | Accenture Interactive. It was a great opportunity to meet and interact with highly intelligent people that love to combine creativity with technology. I always felt like home there. Special thanks to Prakhar, Magno, Thing, Wo, Lisa, Giovanni and Nick for helping me during this project and for making my time in the LAB unforgettable. Special thanks for the people that participated in my interviews, validation sessions and analysis moments. Thanks to all the experts from the Ethics & Philosophy of Technology department of TPM Faculty (gracias totales Juan), the AI researchers from the Computer Vision Lab of the EEMCS Faculty (million thanks to Hayley and Ekin), and all the employees from MOBGEN | Accenture Interactive that helped me during my project.

I want to thank also all my friends who supported me throughout this master and this project. In specific (and in no particular order), I want to thank José Galán for being there when I needed to have some advice about life (and salsa dance), Andrea Montella for forcing me to keep pushing myself and finish my masters, Pip van Esch for your incredibly useful feedback and for being a great listener (and reader), Nien-Hua Gu for all your help during the masters and Honours Programme. Special thanks to all my Latin American (Mexas y no Mexas), Chinese, Taiwanese, German (Melanie :D), and Indian friends from IDE and the TU Delft. I want also thank Miriely Carranza, for being there for me, even when destiny had different plans for both.

Finally, last but not least I want to thank my family for their unconditional support and love. I know that anywhere I'll go, far away or close from home, you will always be there with me. Always. Gracias mamá, gracias papá, gracias hermanito, gracias mami.

Executive Summary

Artificial Intelligence (AI) is one of the most groundbreaking developments of the last half-century. More and more organizations are either implementing AI or considering its application to optimize their lines of business. At the same time, this exponential growth has also raised questions regarding its ethical implications and societal consequences. As many commercial intelligent applications are becoming ubiquitous, and some are getting implemented in higher levels of decision making, it is extremely important to consider the societal and ethical results of this “digital transformation” process. Organizations should understand where their AI comes from as well as whom they are helping or affecting when implementing it.

This situation opens a new window of opportunity for the development of ethical tools that assist companies in the responsible implementation of AI, as well as the current and future business opportunities that this represents. The current thesis explores the ethics of the development of AI applications and its implementation in MOBGEN | Accenture Interactive projects. By including an ethical perspective towards the application of this technology, the organization will improve its value proposition and prevent future unintended and harmful outcomes for the organization and their clients.

Accordingly, the current thesis focuses on understanding the ethical views and future concerns of all the stakeholders involved in the development of AI applications. This is done by following a double diamond design process, which simultaneously includes the execution of a couple of Action Research cycles to generate awareness on the topic inside the organization.

A detailed literature review, several in-depth interviews, and an intense empirical research effort are conducted to understand the current AI ethics context. This reveals a considerable large amount of insights, which are translated to ethical principles and, afterward, transformed into design principles. At the same time, these principles led to the creation of a theoretical framework intended to assist in the ethical evaluation of AI-related projects.

Based on this theoretical framework, a toolkit is designed to help the development teams of MOBGEN | Accenture Interactive with the ethical assessment of AI applications. The toolkit is designed in the form of a “full-day” workshop, which is composed by seven different modules distributed into two major strategic phases. These modules are designed to assist in the generation of ideas and support ethical dialogue creatively and collaboratively.

Finally, a total of six evaluation sessions with MOBGEN | Accenture Interactive and other external stakeholders validated the design toolkit, serving as the perfect preface to recommendations for further development. These recommendations include the development of ethical frameworks using design methods, measuring the ethical impact AI systems, and to enhance research on AI ethics from a design perspective.

Reading Guide

This reading guide explores an overview of the report in order to help the reader find the logic behind the text. A short description of each chapter is shown on the right, disclosing its content and the stage of the design approach where it was conducted. All the chapters of this thesis begin with a little introduction on the topic and the content covered. General conclusions are also discussed at the end of each chapter.

Each chapter's section features a "Discussion" sub-section that is intended to provide a short description of the topics addressed and the learnings gathered. Appendices are attached in a separate section of the report. The names of the people that participated in the empirical studies of the project and the validation sessions, as well as the clients from MOBGEN | Accenture Interactive are anonymized to maintain privacy.

Abbreviations:

AI	Artificial Intelligence
ML	Machine Learning
COBE	Code of Business Ethics
HQ	Headquarters
CV	Computer Vision
EU	European Union
EEMCS	Electrical Engineering, Mathematics & Computer Science
TPM	Technology Policy & Management
IDE	Industrial Design Engineering

Too busy to read? The red boxes resume the general conclusions of each chapter followed by some takeaways for MOBGEN | Accenture Interactive employees. This takeaways are based on practical advices for employees in order to take the most out of the topic as a value proposition for clients.





-  **Discover**
 - 01 Project Context & Approach**
Discusses the context of the project, and defines the problem statement and the research objectives.
 - 02 Artificial What?**
Briefly explores the current state of AI and its future challenges.
-  **Define**
 - 03 The Ethics of AI**
Shares a brief analysis of ethics and its relationship with AI, including its commercial value.
 - 04 Design Research for Context Exploration**
Explores the empirical research efforts made within the organization to test the theoretical assumptions.
 - 05 Transforming insights into design**
The insights gathered previously are transformed into a theoretical framework for AI
-  **Develop**
 - 06 Design Principles for the ethics of AI**
Studies an internal view of MOBGEN | Accenture Interactive and synthesizes the insights gathered for an iterative ideation of ethical solutions.
 - 07 Ethical Toolkit for AI**
Explores the design toolkit created to ethically assess projects that implement AI.
 - 08 Validation of Ethical Toolkit**
Unravels the learnings acquired from testing the different versions of the ethical toolkit.
-  **Deliver**
 - 09 Recommendations, limitations & Implementation**
Recommendations are shared for the implementation of the design tool, as well as current research limitations and future research mplications.
 - 10 Conclusion and personal reflection**
Concludes in a general form and includes a personal reflection.

Table of Contents

INTRODUCTION	9	06 DESIGN PRINCIPLES FOR THE ETHICS OF AI	69
01 PROJECT CONTEXT & APPROACH	11	6.1 Internal Interviews	69
1.1 Context & Background	11	6.2 Generative Exercises & Survey	72
1.2 MOBGEN Accenture Interactive	13	6.3 Action Research Cycle 2 - IMPLEMENT	75
1.3 Project Objective	15	6.4 Strategic Vision of the Project	78
1.4 Research Questions	16	6.5 Iterative Ideation & Toolkit Shaping	80
1.5 Scope of the Project	17	6.6 Chapter Conclusions	85
1.6 Project approach	17	6.7 Takeaways for MOBGEN Accenture Interactive	85
02 ARTIFICIAL WHAT?	21	07 ETHICAL TOOLKIT FOR AI	87
2.1 AI in a nutshell	21	7.1 Strategic Blueprint	87
2.2 AI in the wild (AI applications)	24	7.2 Ethical Alignment	89
2.3 Chapter Conclusions	26	7.3 Project Vision & Values	91
2.4 Takeaways for MOBGEN Accenture Interactive	27	7.4 The AI Application Moral Code	94
03 THE ETHICS OF AI	29	7.5 Chapter Conclusions	95
3.1 A Brief Speech of Ethics	29	08 VALIDATION OF ETHICAL TOOLKIT	97
3.2 Ethics, Technology, & AI	31	8.1 Set-up of Validation	97
3.3 Ethics of AI	32	8.2 External Stakeholder Validation	98
3.4 Ethics' Strategic & Commercial Value	34	8.3 Internal Stakeholder Validation	104
3.5 Chapter Conclusions	37	09 RECOMMENDATIONS, LIMITATIONS & IMPLEMENTATION	113
3.6 Takeaways for MOBGEN Accenture Interactive	37	9.1 Discussion & Implementation recommendations	113
04 DESIGN RESEARCH FOR CONTEXT EXPLORATION	39	9.2 Research limitations & Future research	117
4.1 Research set-up	39	10 CONCLUSION AND PERSONAL REFLECTION	121
4.2 Semi-structured interviews	40	10.1 Conclusion of the Project	121
4.3 Creative Discussion	44	10.2 Personal Reflection	122
4.4 Honours Programme Seminar: Provotype Discussion	47	11 REFERENCES	124
4.5 Action Research Cycle 1 - LEARN	49		
4.6 Chapter Conclusions	50		
4.7 Takeaways for MOBGEN Accenture Interactive	51		
05 TRANSFORMING INSIGHTS INTO DESIGN	53		
5.1 Development Journey of AI apps	53		
5.2 Stakeholder Mapping	55		
5.3 Ethics of AI framework	57		
5.4 The Puzzle of AI Ethics	62		
5.5 Framework & Canvas Validation	64		
5.6 Design Requirements for Toolkit	66		
5.6 Chapter Conclusions	67		
5.7 Takeaways for MOBGEN Accenture Interactive	67		

Introduction

The present project explains the relationship between the topics of ethics and Artificial Intelligence (AI). Ethics, as a philosophical term, refers to standards of right and wrong and has been studied practically since the beginning of civilized human history. On the other hand, Artificial Intelligence is a relatively young technology that is getting more embedded in our day-to-day lives. Its development is transforming the way we see, we speak, and even the way we behave. As this happens, people are raising more and more questions about the ethics associated with its implementation. However, no one knows what would be the consequences in the long run of its fast-paced development. There is only one thing we can be sure of, **Artificial Intelligence is not perfect, therefore, it is dangerously human.**

chapter 1

Project Context & Approach

This chapter provides an overview of the project context and the chosen approach for this thesis based on an extensive analysis on the topic and current situation. First, some background information on AI and its associated ethics is discussed. After this, there is a short introduction to MOBGEN | Accenture Interactive. The chapter concludes with the description of the design assignment, the research questions, and the design approach that was used for this project.

1.1 CONTEXT & BACKGROUND

1.1.1 The Rise of AI

Artificial Intelligence is nowadays one of the most groundbreaking developments of the last half-century. Being a relatively young technology, it has gained a lot of popularity in recent times mostly due to its business potential (Makagon, 2019). Nowadays, more organizations are either implementing AI or considering its application to optimize their lines of business, like monitoring of employee productivity and customer engagement. (Davenport & Ronanki, 2018). It is expected that by 2020 85% of all customer service interaction would be done by an AI-powered digital agent. (Mitchell, 2018) Furthermore, according to Accenture, AI could double world's annual economic growth rates in 2035 by enhancing collaboration between humans and AI. (Accenture, n.d.)

1.1.2 Why Ethics in AI?

As many commercial intelligent applications are becoming ubiquitous and some are getting implemented in higher levels of decision making (i.e. AI judges), it is extremely important to consider the consequences of this "digital transformation" process. The exponential growth that AI

technology has experienced lately not only brings benefits for society, but also has brought concerns about its ethical implications. Labor automation, human profiling, discrimination due to statistical biases, are only some examples of the risks of implementing AI and that, unfortunately, are happening nowadays (Barr, 2015; Levin, 2018). Trust and responsibility are essential for competitive differentiation, which means that organizations cannot take the design of their AI applications slightly.

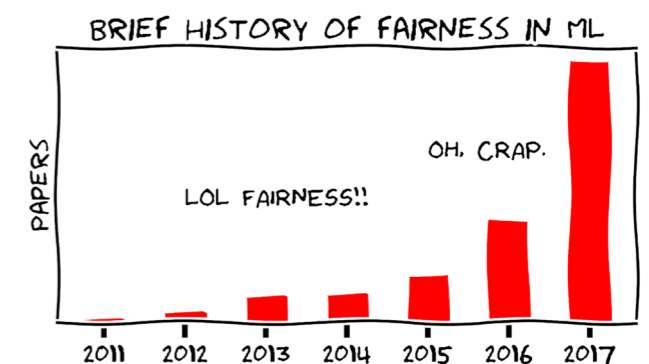
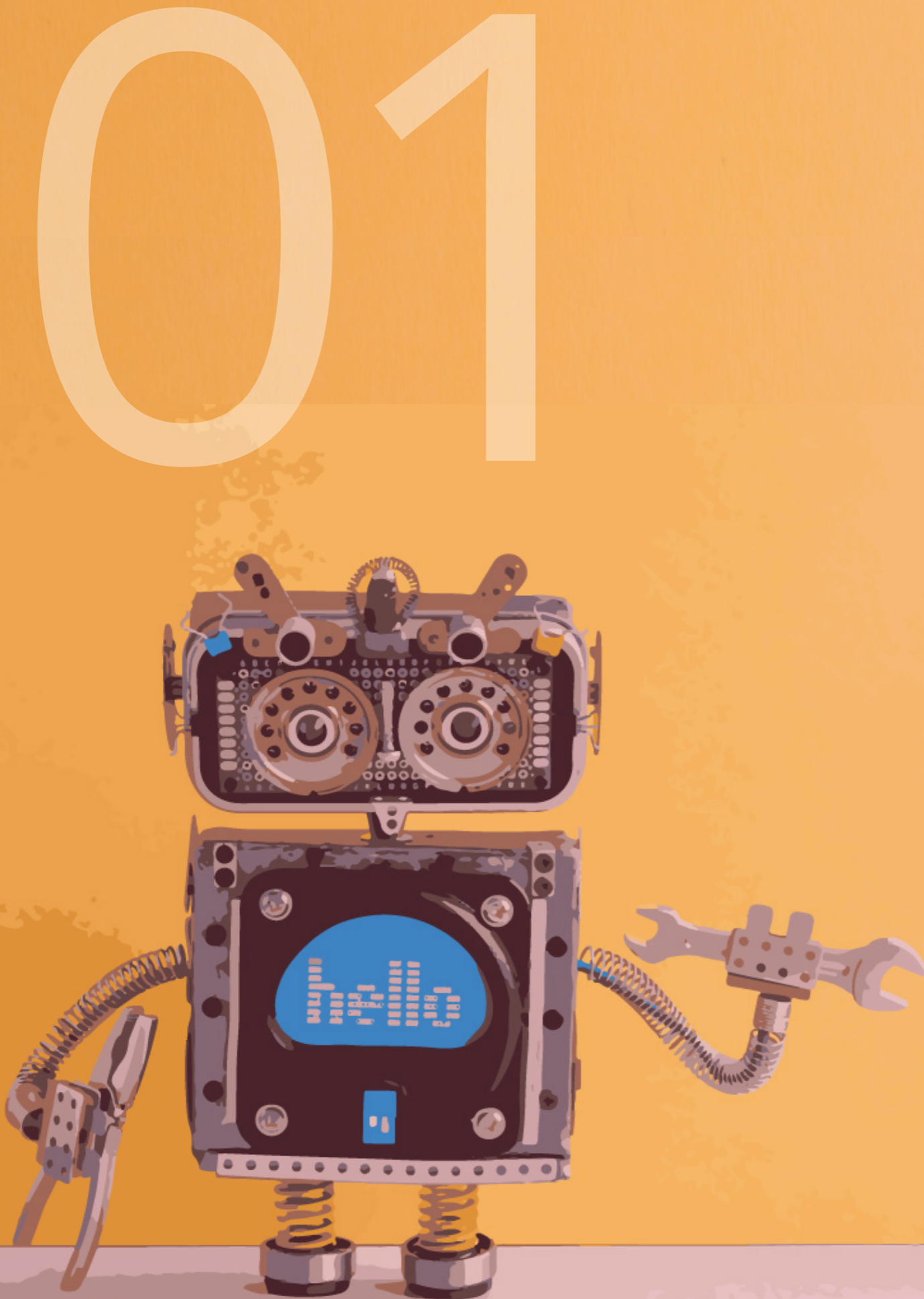


Figure 1: A comical representation of the number of publications on fairness in Machine Learning from 2011 to 2017 (Moritz Hardt, n.d.)



Additionally, Recent academic research findings stipulate that people tend to trust AI applications that simulate a human interaction. According to Robinette (2016), humans will definitely trust robots in an emergency scenario, even when the machines' decision making and judgment are clearly incorrect. Although this might sound quite harmless in principle, we should consider that it is clear that this type of situations bring an enormous ethical responsibility to companies for the correct development and implementation of AI algorithms

Then, it can be concluded that ethics is absolutely relevant for the development of smart applications , although, this is not a new topic within the field. The study of the Ethics of AI has been progressively expanding in recent years within academia as well as in the industry, and it has the potential to help us improve the development of AI technology responsibly (Socher, 2019; Cutler, Pribic & Humphrey, 2018).



Figure 2: Back in 2015, Google's AI confused photos of black people with gorillas (Barr, 2015)

1.1.3 A short overview of the Ethics of Artificial Intelligence

Formally speaking, AI ethics is commonly classified into roboethics and machine ethics. The former is the branch of ethics of technology that concerns the moral of humans as they develop and use AI, meanwhile the latter is interested on the study of the moral behavior of artificially intelligent beings

(Veruggio, 2006). This master thesis focuses only on the roboethics approach as a core insight where design could play an important role. It is important to mention that the ethics of AI is studied using the vast spectrum of philosophy. However, according to Gerdes & Thronton (2015) and based on previous research (Senges et al., 2017), three normative ethical approaches have provided favourable results for the actual implementation of ethical considerations into smart algorithms, a rule based approach called deontological ethics, a future scenario approach referred as consequentialism, and a goodness approach named virtue ethics.



Figure 3: Cover of the draft of the Ethics Guidelines for Trustworthy AI, requested by the EU. (Universidad de Salamanca, 2018).

Nowadays, there is a tendency moving toward solving the ethical behaviour of AI. This means that most of the Ethics of AI research efforts are focused on the development of machine ethics by making, for example, an "Artificial Moral Agent" (Bostrom & Yudkowsky, 2014 ; Dignum, 2018). However, there have been also recent efforts made to inform of the ethical implications that any AI implementation could have. Companies like IBM and Fjord (Accenture Digital) (Cutler, Pribic & Humphrey, 2018; Lubbock & Virdee, 2018) have recently published a series of steps or guidelines aimed at AI developers in order to create awareness of the ethical rules their process should follow. These references include a set of areas of ethical focus like accountability,

value alignment, fairness, and data privacy. More recently, the High-Level Expert Group on AI, which is a group set up by the European Union, published their EU Ethics Guidelines for Trustworthy AI (2019). Within these guidelines, a group of highly skilled experts establish guidelines and parameters for

the ethical development and use of AI. This study involves other ethical considerations like human agency and oversight, technical robustness and safety, and societal and environmental well-being, among others.

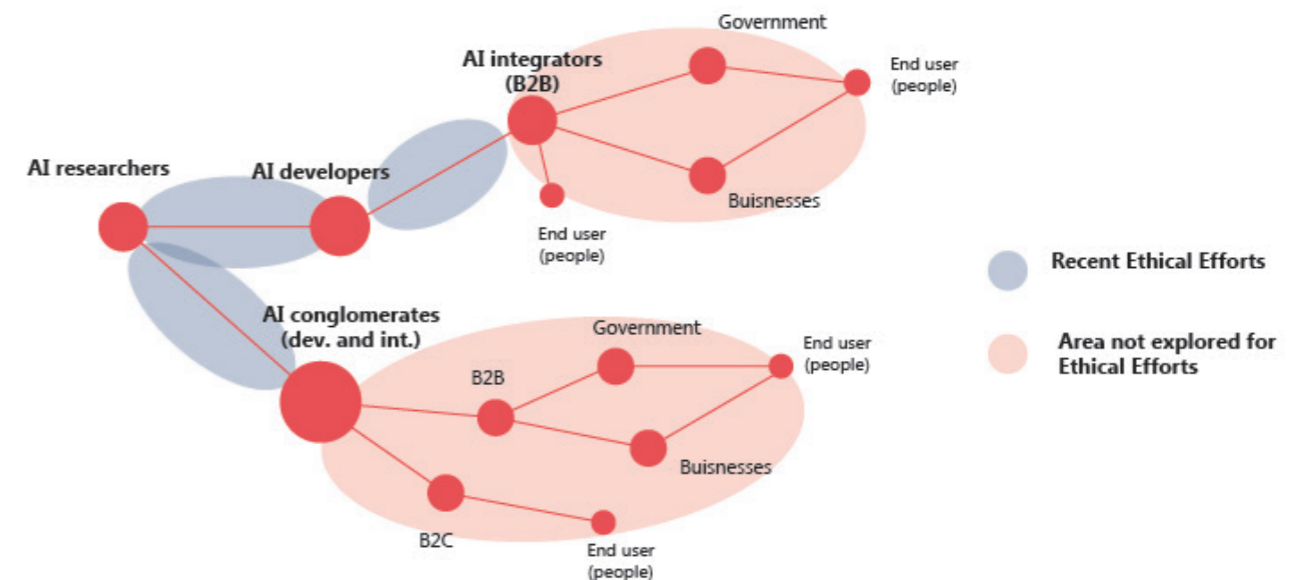


Figure 4: visual representation of the scope addressed in this thesis. The most recent ethical efforts focus only on the research and development stage of AI. Implementors (i.e. MOBGEN | Accenture Interactive) are often overlooked. (illustration by author, 2018)

1.2 MOBGEN | ACCENTURE INTERACTIVE

This thesis is written in collaboration with MOBGEN | Accenture Interactive, however, the design proposed is also intended for other digital creative departments of Accenture like FJORD (The Dock) and Accenture Applied Intelligence. This section describes the context of the project and a summarized analysis of MOBGEN | Accenture Interactive. A general introduction of the company is provided, followed by an overview of Accenture's strategy for its digital branches.

1.2.1 In the beginning

MOBGEN is a digital services consultancy that focuses on the combination of mobile strategies, creativity and technology to develop "one of a kind" solutions to improve the value that brands offer to their clients and partners. Although originated and mainly established in Amsterdam (NL), the company has also offices in Spain (La Coruna and Malaga) and has approximately 400 employees. In 2016, MOBGEN was acquired by the multinational

consultancy Accenture, adding it to its European digital branch with the name MOBGEN | Accenture Interactive. (Accenture, 2016)

1.2.2 Joining Strengths with Accenture

Accenture is a multinational service consultancy with approximately 500.000 employees distributed around the globe (Accenture, 2018), which focuses on providing services like strategy, consulting,

digital, technology and operations. The company's HQ are based in Dublin, Ireland, and its four main strategic pillars are: Consulting, Digital, Strategy, and Technology (Accenture, 2018). These pillars cover a wide range of services from creating digital campaigns, to assessing and changing the ways of working of other organizations. In order to complement and improve its value proposition, Accenture is also open to buy other agencies around the globe, for example, the acquisition of MOBGEN, Storm Digital, Liquid Studio, and FJORD, among others. This thesis concerns a project developed for the Accenture Digital branch, in particular, for MOBGEN | Accenture Interactive.



Figure 5: The former MOBGEN HQs, located in the city of Amsterdam, are currently Accenture Interactive Amsterdam "Building 01". (Dilam, 2016).

1.2.3 Accenture Digital Strategy

Accenture Digital helps clients to discover new growth possibilities by focusing on three areas, Applied Intelligence, Digital Delivery, Industry X.O, and Accenture Interactive. Accenture Digital consultants often execute projects for clients that want to explore the possibilities and opportunities that new technologies bring like Artificial Intelligence, Virtual Reality, Augmented Reality, and Internet of Things. For instance, Applied Intelligence is the part of Accenture Digital that deals with the inclusion of AI at the core of businesses to drive the B2B digital transformation. Furthermore, Industry X.O elaborates upon how companies digitize their core operations transforming their lines of business. In the case of Accenture Interactive, it aims to

provide innovative, end-to-end customer experience solutions with the use of new technologies. This "one of a kind" strategy, that builds upon technology based experiences, brings a clear advantage to the organization and it has been exponentially boosted through the jointure between Accenture Interactive and MOBGEN.

1.2.4 Accenture Values and vision on Responsible AI

Due to the accelerated development of AI, standardized responsible AI practices, which are commonly set by legislations, might take many years to govern this technology. In the meantime, it is the responsibility of organizations to self-regulate its development to ensure that the technology is used responsibly (Chowdhury, 2019). This is the case of Accenture with their recently created Responsible AI team, lead by a diverse class of experts in responsible innovation and AI. The Responsible Artificial Intelligence team of Accenture is an internal group which main objective is to create human-centric Artificial Intelligence. Their main goal is to address the social, regulatory, and economic impact of this technology from all the stages of its development and deployment. The team is currently working on the creation of several tools for the ethical development of AI and have published some successful examples as well (Peters, 2018).



Figure 6: The jointure of MOBGEN and Accenture Interactive represented a change in both organizations (MOBGEN | Accenture Interactive, 2019).

1.2.5 Growth of AI within MOBGEN | Accenture Interactive

In order to create the best customer experiences, the organization has been working to provide solutions that involve in some way a type of AI. From e-stores' chatbots to smart algorithms for an online fuel supplier, MOBGEN | Accenture Interactive clients are embracing the advantages of implementing AI in their applications. Moreover, due to recent consumer trends, it is expected that the need for AI-based experiences increase in the upcoming years (Fjord, 2019). Some of these trends emphasize

1.3 PROJECT OBJECTIVE

Organizations should understand where their AI comes from, as well as whom they are helping or affecting when implementing it. Real-world competitive pressures can cause individuals to make decisions that could have damaging consequences for others. Furthermore, having such a big impact on others' lives enforces the idea that the correct development and employment of AI is a relevant issue. Likewise, there are other implications which are more relevant to design and developing companies, for example, the management of the brand image and brand trustworthiness. This situation opens a new window of opportunity for the development of ethical tools that assist companies in the responsible implementation of AI and also the current business opportunities that this represents.

The objective of this thesis is to create support tools

the need of developing a more responsible AI from the beginning (inclusivity, sustainability), in order to avoid risks and know how to act towards unexpected ethical consequences.

Certainly, one of the goals of this project is to extend the capabilities of MOBGEN | Accenture Interactive in order to support the company when dealing with responsible innovation projects. In this case, with the responsible development of Artificial Intelligence applications.

for MOBGEN | Accenture Interactive to improve their value proposition by including an ethical perspective towards the implementation of AI applications. An ethical tool would allow MOBGEN | Accenture Interactive to prevent unintended and harmful outcomes, which at the same time would protect their own brand and that of their clients in the future. The main reason behind this is that, due to recent ethical scandals around the hightech environment, there is a consumer trust crisis that affects both big and small corporations alike (Francis, 2017; DeMers, 2018). Besides, one of the biggest challenges of the development of technology is the way some professionals fit themselves into a "passive responsibility" state, where a sense of accountability is present only after an unintended event happens (Van de Poel & Royakkers, 2011). With the addition of ethical tools to their current offer, MOBGEN | Accenture Interactive would also help

PROJECT DESIGN ASSIGNMENT



Design **ethical support tools** for MOBGEN | Accenture Interactive to **improve their value proposition** by **including an ethical perspective** towards the implementation of AI applications.

their clients to expand their use of AI and at the same time allow them to be aware of their accountability and the unintended ethical consequences that this application could produce. Subsequently, the outcomes of the collaborations with its clients will be more ethical and socially desired.

It is important to mention that the main line of business of the company is to create digital experiences for customers, which could include the creation of mobile apps, new technologies exploration (i.e. mixed reality) or service design sprints. Hence, the implementation of AI comes only as an added benefit for their clients and not as a main value proposition. This could be scaled to the tools created as its purpose would be to trigger and support dialogue with clients regarding the topic of ethics of AI. This approach would support the innovation culture at MOBGEN | Accenture Interactive by providing a critical framework to

address the implementation of AI.

The objective of this project was achieved by an approach that mixes academic and professional techniques which will be described in more depth later on.



Figure 7: The “Artificial Mirror”, an experience made by MOBGEN | Accenture Interactive for explaining the training mechanisms of AI (Mobgen, 2019)

1.4 RESEARCH QUESTIONS

Currently, there is a need for more integral ethical efforts regarding the development and implementation of AI applications. In order to achieve this, the following research questions guided the project:



Q1: What are the ethical views and future concerns of all the stakeholders involved in the development of AI applications?

Q1.2: How does this knowledge help in producing an operational framework for AI ethics?



Q2: How to support MOBGEN | Accenture Interactive with tools for a responsible development of AI applications?



Q3: What are the strategies that MOBGEN | Accenture Interactive could follow to provide an added value proposition to their customers through the ethical uptake of AI applications?

1.5 SCOPE OF THE PROJECT

1.5.1 Project Extent

This project was performed in collaboration with MOBGEN | Accenture Interactive. All the tools designed during this thesis are made considering the context of the organization as well as the context of Accenture in general. This includes the collaboration of other Accenture Digital dependencies like FJORD and Accenture Applied Intelligence. Furthermore, this thesis only explores AI from a high level perspective in order to support the organization with the responsible development of the technology and its applications. The proper development of smart algorithms and other AI techniques are not considered or explored in detail in the present project.

Additionally, as ethics is a broad philosophical discipline, it is important to mention that this project only explores the application of three

ethical branches (consequentialist, deontological, virtue) with a pragmatic point of view, within the context of AI.

1.5.2 Involved Stakeholders

The parties involved on the development of this project include Ethics of Technology researchers (TU Delft), AI researchers (TU Delft), AI implementing experts from MOBGEN | Accenture Interactive, AI enthusiasts, and AI users, among others. Additional stakeholders like experts in the field of Internet of Things and Design with Data supported the research and validation stages of the project. Other parties involved incorporate the TU Delft supervisory team and the company mentor.

1.6 PROJECT APPROACH

1.6.1 Double Diamond design method with a twist

In the broadest sense, this project follows a typical double-diamond process (Design Council, 2005) which is a representation of a double loop of diverging and converging stages. The process is composed by four main stages which are: discover, define, develop, and deliver. Furthermore, both diamonds have been enriched with elements of “action research” as shown in Figure 8 (Price et al., 2018). Action research is a research process that aims to transform based on taking action and doing research simultaneously by linking this together with a reflection process (Reason & Bradbury, 2001). One of the reasons for the application of action research was the intention of transforming the organization’s and public perception around the ethics of AI, as this is often taken more as an “obstacle” in the way instead of as a valuable tool for innovation.

Additionally, an extra method has been considered during the “Discover” stage of the process. This process involves the application of normative ethics approaches towards the investigation and understanding of the current ethical AI environment. By inquiring about the possible ethical failures, important ethical insights could be collected from the empirical research. These insights would be related to the unintended consequences of AI applications. The project approach process is briefly described next:

Discover

The main focus of this stage was to understand the Ethics of AI and the way it is perceived by the main stakeholders involved in the development process of AI applications. This was achieved through an extensive literature review as well as by performing several interviews (and generative exercises) with relevant stakeholders (researchers, implementors, users). The interviews were performed taking into consideration a deontological and consequentialist ethical approaches, in order to convey in more robust results. The results from this phase led to a more established design direction.

Define

The main goal of this phase was to converge the design direction towards a more concrete solution space. Several design techniques were applied to create a more concrete background of the problem by converging upon the insights acquired in previous stages. A presentation of the topic through an academic seminar helped to complement these insights. Furthermore, continuous interviews with experts and employees helped to find existing solutions and the preferred design direction that the organization would like to explore regarding the Ethics of AI. This phase ended with the definition, embodiment, and expert validation of an ethical framework for the development of AI applications.

Develop

In order to make the insights around the ethical framework for AI applicable for MOBGEN | Accenture Interactive employees, design opportunities were explored to get to a final design solution. Several expert interviews and a couple of generative exercises helped to find the right approach towards the preferred design direction of MOBGEN | Accenture Interactive. Moreover, some “provotypes” (action-research) were used to create awareness around the importance of the Ethics of AI within the company. Finally, the ideated design solutions were iterated multiple times with the support from all the stakeholders involved during the previous phases. A couple of workshops runs were evaluated as well, validating the impact and functionality of the design solutions.

Deliver

In this phase a final design was created based on the information compiled on previous stages and the received recommendations. An implementation model taking into account the processes of the company was included as well. Several final tests with internal and external stakeholders were conducted to validate the design.

Design Methods used

Multiple design methods were used during the four different phases of the project. These include a desk research, several semi-structured interviews, persona generations, creative sessions, prototyping and provotyping as well.

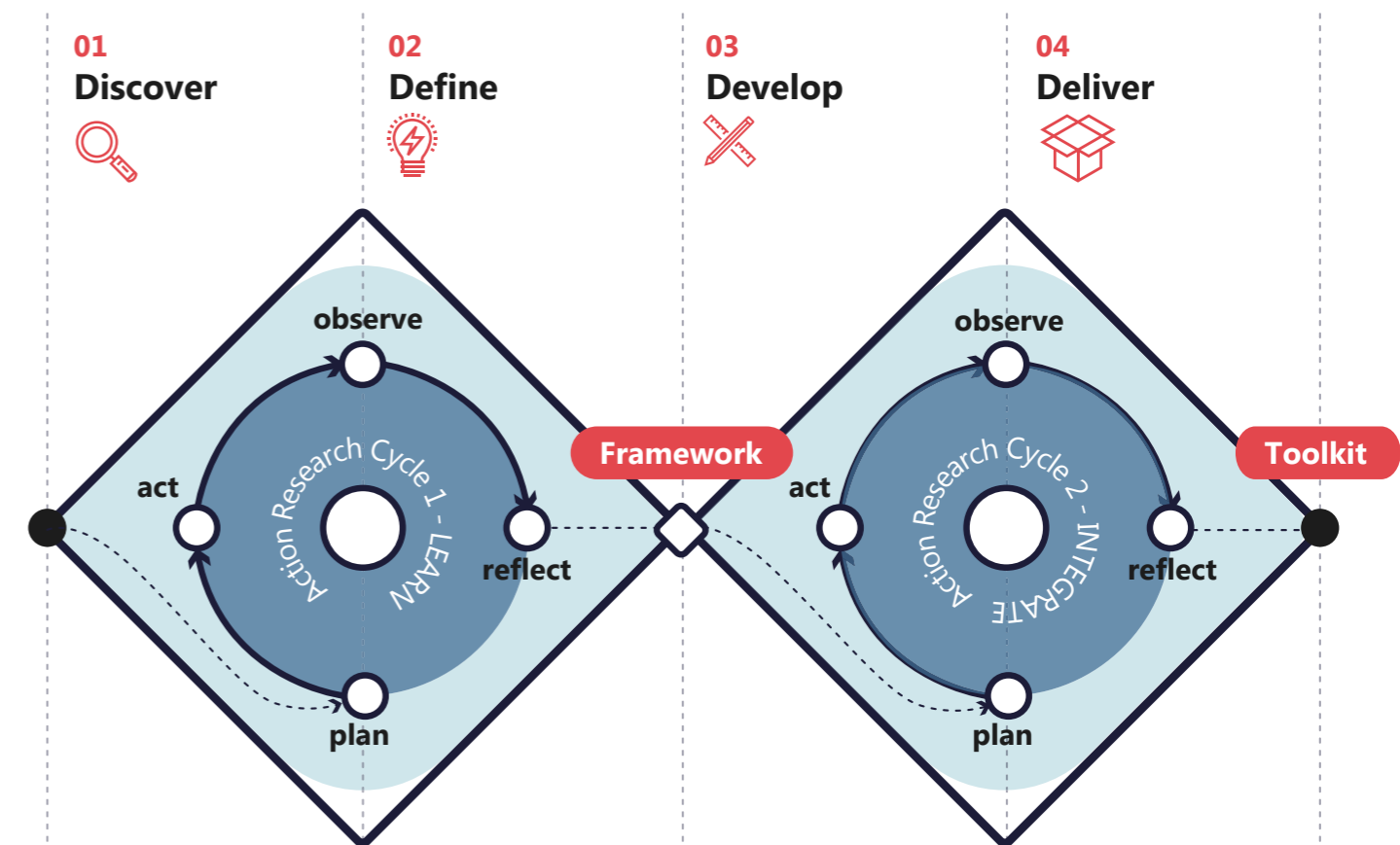


Figure 8: Visual representation of the approach followed during this project following a couple of Action Research cycles (illustration by author, 2019)

chapter 2

Artificial what?

This chapter explores the definition of Artificial intelligence, as well as its current capabilities and future challenges. Due to its recent increase in popularity, a lot has been written about the topic in the media, academic research, and even in other master theses. Hence, this chapter covers only a brief overview of AI and its current state. Furthermore, this chapter also includes the definition of an AI application, the stakeholders involved in its development as well as a rough development journey visualization.

2.1 AI IN A NUTSHELL

This section briefly explores the definition of Artificial Intelligence, its current capabilities, and the challenges that the field faces towards the future.

2.1.1 What is exactly AI?

In computer science, AI is defined as any type of intelligence displayed by a machine. Its research focuses on the study of “intelligent agents”, which at the same time are defined as anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators (Russell & Norving, 2009). This means that AI could be found in tangible objects (i.e. smart gadgets, robots) as regularly as on “intangible” entities (i.e. stock market, Google’s AlphaGo). According to Russel & Norving (2009), the main goals for the development of an AI agent are that it should be capable of interaction with the environment, keeping track of the state of the world, evaluating and selecting the future courses of action, and finally learning from the previous cycle of knowledge.

Although people tend to understand AI as a “single type of technology”, the AI present in a recommendation engine is not the same as the one found in an Intelligent Personal Assistant like Alexa



Figure 9: AphaGo is a great example of a narrow AI which was able to learn “Go” and beat the human world champion Lee Sedol (Fingas, 2016).



or Google Home. It is important to mention that Artificial Intelligence is an extensive discipline that is composed of a diversity of paradigms. Broadly speaking, this technology could be classified as narrow and general AI (or Computational Intelligence and Symbolic Artificial Intelligence paradigms). The former focuses on methods that simulate specific human cognitive abilities. Some of these applications include perception and pattern recognition, knowledge representation, problem-solving, reasoning, decision making, amongst others. Furthermore, this paradigm is responsible for the development of some popular AI techniques, like machine learning and deep learning, which have been implemented in almost all smart device and application nowadays (Flasiński, 2016). On the other hand, general AI concentrates in the replication of human intelligence as a whole, not only on specific cognitive tasks. This paradigm is still in an initial stage since its development has proved to be more complex and less commercially reliable than narrow AI.

Because of the reasons above, and due to popular understanding, it is useful to define the umbrella term of Artificial Intelligence as exclusively the group of specific paradigms designed to perform specific tasks. Therefore, in the particular case of this thesis project, the term “AI” will be used to refer to the narrow Artificial Intelligence technologies (i.e. machine learning, deep learning, etc).

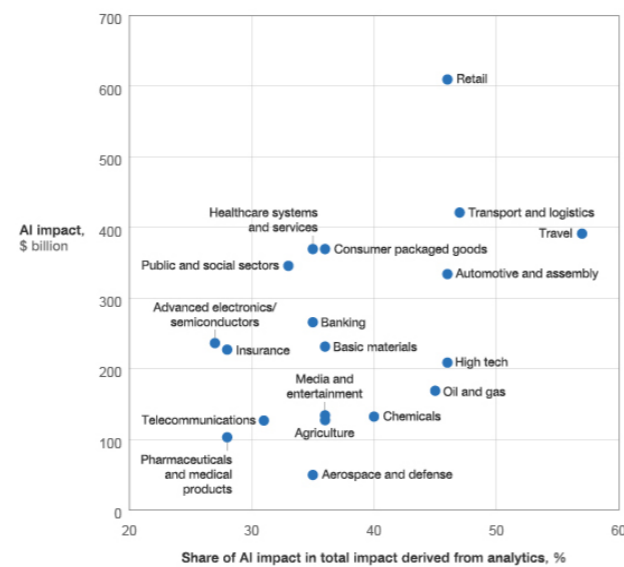
2.1.2 The State of AI Today

Artificial Intelligence (AI) is a relatively young technology (been in development since the 1950's) which has gained a lot of popularity in recent times, mostly due to its business potential (Chui et al, 2018; Purdy & Daugherty, 2018). This technology has the potential to create a high amount of value for different commercial sectors, from Retail to Pharmaceuticals as shown in Figure 10.

Up to 2019, AI is capable of doing things that were not possible a decade ago. Some of AI's current capabilities englobe the ability to sense and

recognize certain types of input signals such as images, sounds and text, as well as the ability of recognizing patterns and predict actions based on this input (Noessel, 2017). Moreover, according to Davenport & Ronanki (2018) companies believe that the actual capabilities of AI would support some of their operational activities like optimizing internal business operations, automating tasks and assistance to make better decisions.

Artificial intelligence (AI) has the potential to create value across sectors.



McKinsey&Company | Source: McKinsey Global Institute analysis

Figure 10: AI impact across different sectors according to McKinsey & Co. (Chui et al., 2018)

2.1.3 The Challenges around AI

The exponential development that AI has experienced in recent years carries important challenges as well. Some of the most referred challenges in the literature are presented next:

01 Data still leads

Something that is important to consider is that AI is a technology that requires of massive amounts of data to work properly (Flasinski, 2016). This limitation keeps the development of AI in a slow pace as the data required has also issues related to preprocessing methods and ethics (McKinsey, 2018)

02 High expectations

Mainly due to the heavy media coverage that AI has been experiencing lately, the expectations that people have around the field are one of the big barriers in its development (Chen & Asch, 2017; Burgess, 2017)

03 Automation and oversimplification

Several authors have pointed out the impact that AI will have on jobs as part of the biggest barriers of its implementation (Burgess, 2017; Mckinsey, 2018; Reisinger, 2019; Accenture, n.d.). Whether the impact of the implementation would be positive or negative is still in debate, nevertheless, several organizations

have started to automate their operations and processes.

04 Lack of diversity in the field

The social diversity of most of the current AI development teams is low, which means that the decisions made during the development process do not consider different ethical points of view and therefore do not represent a wide overview of social values, which would enforce historical biases and power imbalances (Chekanov et al., 2017; Paul, 2019).

05 Ethics of AI

As mentioned in several sources, the ethics involved in the development of AI is considered an important challenge the AI field face today and will keep facing in the future. (Belei, 2019;McKinsey, 2018).

2.1.4 Discussion

A lot has been published recently about the current development of AI and its challenges for the future. Although, we have not reached a point of having developed “true” Artificial Intelligence, it has been proved that its “narrow” applications have the potential to generate commercial value and create an important societal impact. Nevertheless, the exponential development of the field has brought important challenges as well to consider, mostly regarding the human side of the technology. For example, data is still necessary for the correct development of AI systems, which brings challenges on its own. Moreover, the high expectations that have been generated around AI creates an unrealistic picture of the technology. Also, lots of doubts have been raised regarding the ethical issues that this technology brings like the lack of diversity on the development teams and the unavoidable automation AI will provoke. This section covered a brief overview of the current state of the technology, as a more particular case is discussed in the next section: AI incorporated into commercial applications.

2.2 AI IN THE WILD

The focus of this thesis turns around the idea of a commercial AI application and the ethics involved in its creation. For practical means, an AI application is defined as any AI model or algorithm that is implemented in a commercial application or product. This section describes the current stakeholders involved in the development of AI applications, as well as a “high level” version of the development process of an AI application.

2.2.1 An AI Application

An AI application could be defined as any AI model or algorithm that is implemented into a product or digital commercial application that would be used by the general public (B2C). Some examples of these include customer service bots, Intelligent Personal Assistants (i.e. Siri, Alexa, Google Assistant), or the embedded AI algorithms in applications like Google Maps, Tinder, and Uber, among others (Polachowska, 2019). As mentioned in Chapter 1, several efforts have been made in order to implement an ethical assessment specifically to the research and development of AI systems done by high-tech companies and research centers like Google Brain and IBM (Simons, 2019). However, it is not the case when we talk about the implementation of this technologies in commercial applications, where a lack of “double checks” might be present and that could be represent a risk in the future.

Certainly, in order for an AI algorithm to get into a commercial application, it has to go through a long development journey which starts with its creators, generally data and computer scientists, and ends up with the general user of the application. The stakeholders involved in the process of an AI application are discussed next, as well as a rough estimation of the development journey of an AI application.

2.2.2 Stakeholders Involved in the development of an AI application

From a general perspective, the main stakeholders involved in the value chain development of an AI applications are AI developers, AI user companies, and AI users (PDPC, 2018). The “AI developers” group include the developers of state-of-the-art AI models, which for this project are referred to Researchers, and also the developers that integrate AI-powered features into products, referred as Implementors.

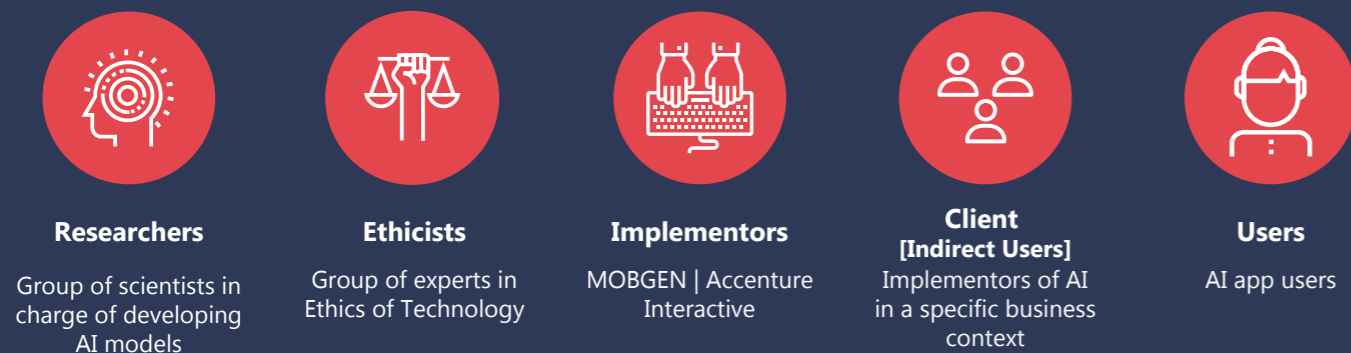


Figure 11: A simplified overview of the stakeholders involved in the development of an AI application (illustration by author, 2019).

Meanwhile, the group of the “AI user companies” is composed by those companies that make use of AI solutions for their products, services, or implemented into their operations. For the purpose of this project, this group would be referred as Clients or Indirect Users. Something that is worth to mention is the special role played by experts in ethics or Ethicists (included in the “AI developers” group) as they have been incorporated recently into the development cycles of AI (Mckinsey, 2019). As ethics becomes more important for the field of AI, it is expected that this role would have a bigger presence in the future. A visual overview of the stakeholders is shown in Figure 11.

As different stakeholders have different roles, they occupy a different level in the value chain of the development of AI application. A stakeholder map featuring an “onion” model is shown in Figure 12 in order to explain this importance. The core of the chain is composed by the Researchers as well as the Ethicists as they have the biggest responsibility in the development process. The middle layer includes the Implementors and the Clients, as the final layer is composed by the Users of the AI application as they acquire all the value that comes from the development process.

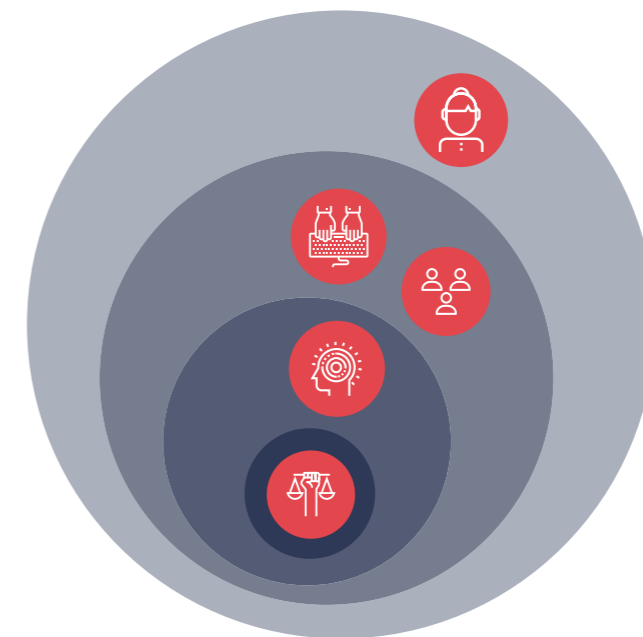


Figure 12: Stakeholders map created as an “onion model” where the involved stakeholders are classified into levels depending on the importance of their role (more important = core) within the development of AI applications (illustration by author, 2019).

2.2.3 The Journey map of an AI Application

A very simple and concrete journey map of an AI app development is shown in Figure 13. The process begins with the research stage performed mostly by the AI researchers where a state-of-the-art AI technique is discussed (i.e. A CV algorithm using a Convolutional Neural Network to identify text). The next step begins when companies like MOBGEN | Accenture Interactive compile the created algorithms or use the algorithms already processed by a third party (i.e. Microsoft) and they implement

it in an application for a client. This client provides a B2C model and has selected MOBGEN | Accenture Interactive to create an application that would use any type of AI to satisfy a user need. The current role of the experts in ethics or ethicists is quite limited and it is mainly focused on the core stage when the AI algorithm is developed. It is important to mention that the activity followed by each stakeholder has its own process which generally involves, data, training of algorithms and model generation.

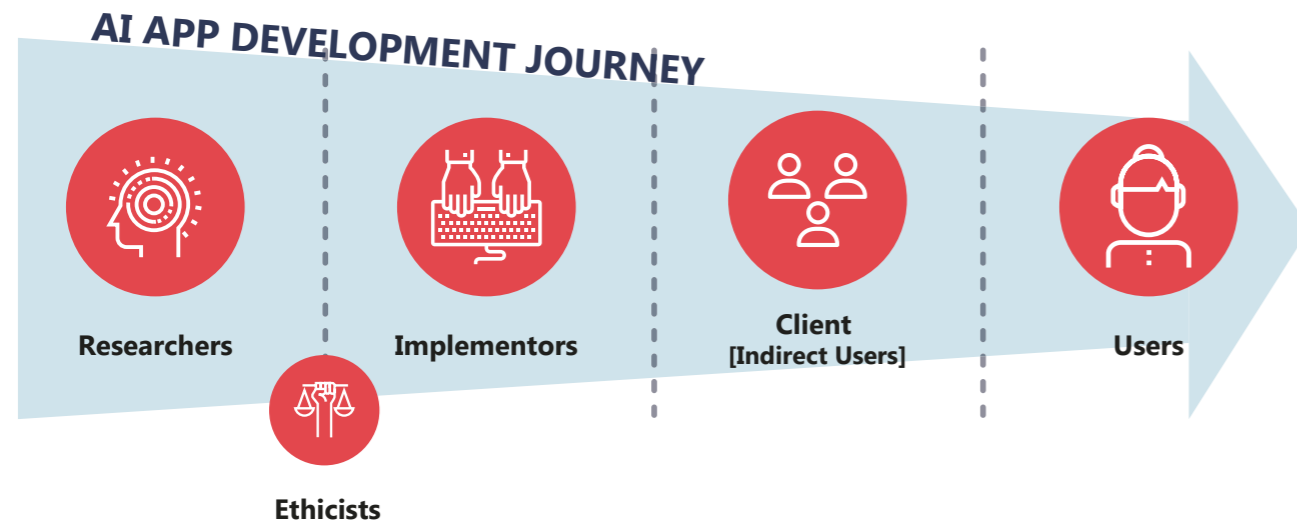


Figure 13: Visual representation of a rough Journey map of an AI application (illustration by author, 2019).

2.3 CHAPTER CONCLUSIONS

A lot has been written about AI during the last couple of years due to a grow in popularity about the topic. Therefore, only a basic description of Artificial Intelligence is provided in this chapter. For practical purposes, the term “AI application” is defined in this project as any AI model or algorithm that is implemented into a product or digital commercial application that would be used by the general public. Some examples of these include customer service bots, Intelligent Personal Assistants, and AI models implemented in digital platforms like Uber or Tinder. Additionally, the stakeholders involved in the development of AI applications include AI Researchers, Implementors, Clients or Indirect Users, and Final Users. As studied in this chapter, the field of study of AI has been experiencing some great developments around its “narrow” paradigm, in the form of sensing and recognizing images and sounds, as well as in pattern recognition and output prediction. However, this also brings several challenges mostly concerning the human aspects of this technology, for example, with the ethics involved on its development and implementation, as discussed in the next chapter.

2.4 TAKEAWAYS FOR MOBGEN | ACCENTURE INTERACTIVE



MOBGEN | Accenture Interactive designers **should take time to establish a strong narrative** around the AI solutions they could provide that also provides a realistic view on what the technology is capable of doing today.



It is **key to identify and clearly communicate the challenges around AI** and where does the company and its clients fall within the development journey of an AI application. This would help the company to establish a trustworthy relationship.



As AI is currently a fuzzi-word on the media and pop culture, it is **important to understand that clients might not have a lot of previous experience with the technology**. Hence, it is important to identify their knowledge and steer it towards a narrative that fits the project and the responsible innovation view.

03

chapter 3

The Ethics of AI

This chapter explores an overview of Ethics, the Ethics of Technology, and the Ethics of AI. Also, an overview of the business ethics followed at MOBGEN | Accenture Interactive is addressed. Moreover, the final section describe the commercial value of ethics.

3.1 A BRIEF SPEECH OF ETHICS

This section delves into the main insights of ethics found by doing a literature review and interviews with experts. To avoid falling into a philosophical loop, a pragmatic lense will be used as ethics is described and defined in a practical way using concrete examples.

3.1.1 What is ethics?

Defining ethics is a complex task as in the majority of the cases trying to do so stimulates a philosophical debate around the topic. A very common misconception is to define ethics as simply the relationship between what is right or wrong. This definition, that assimilates more to the definition of morals, lacks of the depth of context that ethics proposes. Properly, ethics (also known as moral philosophy) involves the systematizing, argumentation, and defending of the concepts of right and wrong behavior. (Internet Encyclopedia of Philosophy, n.d) Nowadays, ethical theories are divided into three general sub-areas: metaethics, normative ethics, and applied ethics. Metaethics investigates the origin of our ethical principles, normative ethics are in charge of propose moral standards that regulate our conduct. Furthermore, applied ethics focuses more on the study of specific ethically controversial issues, for example, abortion, animal rights, homosexuality, etc. (Internet Encyclopedia of Philosophy, n.d)

Although, the differences between ethical theories are often blurry, the most popular approach around the ethical development of AI deals with normative ethics (Gerdes & Thronton, 2015; Seneges et al.,

2017). The three most important normative ethics approaches are described in Figure 14.

 Normative Ethics Approaches	
Consequentialist	<i>The best outcome for the largest amount of people</i>
Deontological	<i>Doing the right thing and performing the right action (no matter what, you should follow the rules)</i>
Virtue	<i>Being a virtuous person</i>

Figure 14: The three most important normative ethics approaches (illustration by author, 2019).



3.1.2 Ethical decisions

Ethics is based on making decisions towards different types of situations. According to Zhou (2018) and as shown in Figure 15, we find three types of potential ethical situations in our everyday lives, these are everyday situations, accidents, and dilemmas. The latter is an extreme case that perhaps we would never experience but it works as an experiment to test our morals (i.e. Trolley Problem ethical dilemma). On the other hand, ethical accidents are something that can happen and that require our own ethical considerations to analyze the situation and take action. These situations require of ethical judgement, which comes from the personal inclination towards any of the previously defined normative ethical approaches (deontological, consequentialist, virtue). Hence, it can be stated that when facing an ethical situation we tend to use an ethical framework to try to make a concrete decision.

In the case of the development of AI applications, it has been discussed in previous research that designers and engineers tend to design and develop with their own morals and ethical approaches on the line (Vincent, 2019). Furthermore, it has been also stated that the social diversity of most of the AI development teams is low, which means that the decisions made during the development process do not consider different ethical points of view

(Chekanov et al., 2017; Shilton, 2018). A high diverse environment, which includes the participation of women, minorities, people of different ethnicities, and people with disabilities, would avoid the spark of historical biases and power imbalances (Paul, 2019). Hence, this would mean that a high diverse team could develop an AI system with minimal ethical issues as they would include more discussion into ethical decisions. This has also been previously studied when the integration of direct and indirect stakeholders in the early stages of the ethical decision making could bring benefits to the process (Van de Poel & Royakkers, 2011; Mephram & Kaiser et al., 2006).

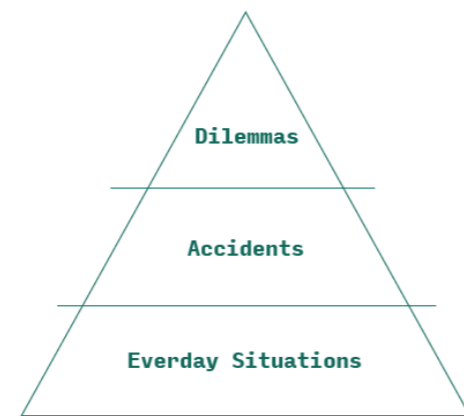


Figure 15: Three different ethical situations we might face everyday (Zhou, 2018)

3.1.3 Discussion

To conclude, it is common that people face situations where they would have to make a decision that involves their ethical views. As mentioned before, in the case of the developers of AI systems, they tend to use their own morals for the development of the technology, which of course includes the ethical decision making (if any) that happens during the process. This is something that could cause ethical harm as it is well known that the people in charge of the development of AI belongs to a highly homogeneous environment. Hence, the process of ethical decision making in AI development could be benefited by executing actions that encourage the diversification of the field and the inclusion of direct and indirect stakeholders in early stages of the development process. To complement this, it is important to understand the different approaches that ethics cover within the development of technology as explained next.

3.2 ETHICS, TECHNOLOGY, & AI

This section briefly discusses the contemporary fields of Ethics of Technology and Responsible Innovation, and how it is addressed within organizations.

3.2.1 Ethical approaches to technology

The ethics of technology is the sub-area of ethics that controls and tries to answer the ethical questions of the development of technology. Also known as technoethics, its objective is to create theories and methods to ethically evaluate technological systems and practices. This area of ethics promotes the ethical use of technology as well as protects against the misuse of it (Luppardini, 2010). Additionally, technoethics is closely related to Responsible Innovation, which is introduced in academic literature as an actual framework that allows the satisfactory uptake of moral issues in technology development (Pesch & Werker, 2017). Moreover, the European Union has pushed the term in order to make societal actors work together to align the innovation results with societal values and needs (European Commission, n.d.).

An important example of the application of the ethics of technology and the responsible innovation approaches is found in the current development of AI systems. As its developers are exploring AI to understand the world around and to make our lives more efficient, it is fundamental that they understand the societal implications of the consequences. This is because, among other things, it does not mean that only because something is more efficient it is morally better for society (Green, n.d.). As van de Poel correctly states (2011), “engineers are supposed not only to carry out their work competently and skillfully but also to be aware of the broader ethical and social implications of engineering and to be able to reflect on these”.

3.2.2 Codes of Conduct: Possibilities and Limitations

A code of conduct is a code made by organizations that feature the behavior guidelines expected from

their members. These guidelines should guide the values that would guide behavior and the decision making process of the organization (Van de Poel, 2011). Codes of conduct are often made as a complement to legal requirements as they are mainly enforced by the organization that created them. There are two types of codes of conduct that are especially important for technologists: professional codes and corporate codes. The first one refers to codes of conduct formulated by professional associations that inform and support technologists with best practices and knowledge update. The second ones are the codes of conduct made by the companies that employs technologists.

The codes of ethical conduct are not a “one-stop-store” solution to ethics in organizations as they are currently pretty limited (Dobson, 2003; Webley & Werner, 2008; Norberg, 2009). Because of this, several authors have determined that an ethical culture based on a promoted ethical program is necessary to ensure an ethical behavior by professionals (Webley & Werner, 2008; Beeri et al., 2013). An ethical culture generally follows a top-bottom approach, which starts with actions from top management or leadership people. We cannot expect that other stakeholders behave in an ethical manner if they perceive that they have been guided in a dishonest or unfair manner, or if they are not allowed to point out and discuss about unethical behavior. Likewise, an ethical program would involve a code of ethics, a proper guidance, an efficient and secure system of ethical advice and protection, and an ethical training (McMillan, 2012).

3.2.3 Discussion

This section tries to briefly describe some ethical approaches that are taken for the development of technology and its development. An important insight that can be concluded from this section is that most of the theory behind the Ethics of Technology is currently too theoretical as most of the approaches rely on theoretical frameworks to try to regulate the behavior of technologists. On the other hand, to follow an approach of good practices, several organizations have conceived codes of ethics in an attempt of making a tangible set of behavior regulations. It is important to mention that it is wrong to assume that implementing a code of ethics will make people align their behavior to the code as several authors have demonstrated. In order for this to happen, a more integral ethical strategy should be taken, which would contemplate the development of an ethical culture and the implementation of an ethics program. This is more relevant when we talk about the recently created field of AI ethics, which is explored in the next section, and which is experiencing a considerable big boost of popularity.

3.3 ETHICS OF AI

This section elaborates on the concepts related to the Ethics of AI and its current state of development today. This is important since it shows where companies are standing in the topic of the ethics of AI, an overview of the public opinion, and the current efforts that are being performed in order to prepare developers and designers for the ethical challenges of AI

3.3.1 Ethics of AI & the public opinion

As mentioned in Chapter 1, ethics of AI is studied using the vast spectrum of philosophy. However, for practical matters this thesis focuses on roboethics which is the branch of ethics of technology that concerns the moral of humans as they develop and use AI. This is relevant due to the recent fuzzi and interest that the topic has triggered on people, mostly by the media and the open concerns of policy makers. For instance, the chairman of the Dutch Data Protection Authority (AP) expressed his opinion regarding the role that the government has on the development of smart algorithms. He demands that the development of AI should be transparent about the use of predictive algorithms as Citizens also have the right to know how those calculation models come to a conclusion (Van der Beek, 2019).

Additionally, the topic of ethics of AI was addressed during this year's Bilderberg meeting (Bilderberg, 2019), which shows that the topic is not considered a niche subject anymore by the western free market potencies. Furthermore, the agenda of the World

Economic Forum at Davos 2019 featured several talks and round tables on the topic (Tyagarajan, 2019). For instance, a discussion session was organized with the topic "Is Responsible AI Good Business?" where various experts in the field argued about the business potential of ethics within the development of AI. According to the discussion, ethics will become a strategic competitive differentiator for companies in the future as people will look for trustworthy sources of products and services.



Figure 16: Accenture's panel at the World Economic Forum with the discussion topic "Is Responsible AI Good Business?" (Accenture, 2019b)

3.3.2 Lessons learned from faulty AI implementations

Although the implementation of AI has become popular in recent years, several ethical faulty implementations have provoked the recent interest in the topic. For example, back in 2015, users noticed that Google's photo app, which uses AI to automatically classify pictures, was labeling images of dark-skinned people as gorillas (Barr, 2015). Another example that is important to consider is Microsoft's Twitter bot called Tay, which learned to "tweet" by interacting with people's opinion found on this popular social network. Unfortunately, due to the nature of this biased data (toxic points of view from some users), the system started behaving in a racist manner and was taken offline shortly after (Victor, 2016). Even natural language processing techniques like the ones used on smart assistants like Alexa and Siri are prone to ethically fail, like the case of Amazon's Alexa recommending adult content to a little child (Entrepreneur, 2017).



Figure 17: "Tay", a twitter bot developed by Microsoft, started to behave unethically shortly after being released online. (Hope, 2016)

All these examples share a common characteristic, which is that regardless the type of implementation and application AI could have, it is important to reflect in advance all the possible cases in which the technology is not aligned the with values and ethical principles of a society or community it affects.

Everyday Ethics for Artificial Intelligence



Figure 18: IBM's Everyday Ethics for AI (Cutler et al., 2018)

3.3.3 Ethical Guidelines for a responsible development of AI

As mentioned previously, there have been recent efforts made to inform of the ethical implications that the development and implementation of AI could have. Companies like IBM and Fjord (Accenture Digital) (Cutler, Pribic & Humphrey, 2018; Lubbock & Virdee, 2018) have recently published a series of steps or guidelines aimed at AI developers in order to create awareness of the ethical rules their process should follow. These references include a set of areas of ethical focus like accountability, value alignment, fairness, and data privacy. This topic is also addressed by Google in its "People + AI" guidelines (2019). More recently, the High-Level Expert Group on AI which is a group set up by the European Union, published their EU Ethics Guidelines for Trustworthy AI (2019). Within this guidelines, a group of highly skilled experts establish guidelines and parameters for the ethical development and use of AI. This study involves other ethical considerations like human agency and oversight, technical robustness and safety, and societal and environmental well-being, among others.

3.3.4 Discussion

Referring back to the content of this section, it can be noticed that a lot more work needs to be done in the field of Ethics of AI. During the last two years there has been an increasing amount of events, foundations, and collaborations organized around this topic. As mentioned previously, companies, research institutes, and policy makers are currently publishing ethical visions and principles around AI, in the form of guidelines, to advise responsibly to the implementers and developers of this technology. Nevertheless, due to the complexity of the subject, little practical methods on how to trigger and even implement this ethical principles during the development of AI have been addressed.

3.4 ETHICS' STRATEGIC & COMMERCIAL VALUE

This section explores the field of ethics within a business context. It also delves into how organizations like MOBGEN | Accenture Interactive address it and its future commercial value.

3.4.1 Business Ethics

Business ethics is a form of applied ethics that investigates the ethical principles and ethical issues that can arise within a business environment. (Moriarty, 2017) It also rules all the aspects of business conduct of individuals and organizations. It is generally embodied in organizational standards, values, and norms (codes of ethics) that control the behavior of individuals in the organization, as well as the actions of the organization itself. (Van de Poel & Royakkers, 2011)

Although, it has been found that ethics is often ignored in the business world due to its complexity (Davis and Patterson, 2012), most of the major corporations today promote a non-commercial commitment to moral values under codes of ethics and social responsibility campaigns. These codes of ethics allow organizations to give access to difficult ethical concepts and promote an ethical behavior among their employees.

3.4.2 MOBGEN | Accenture Interactive Code of Business Ethics

MOBGEN | Accenture Interactive follows the same code of business ethics as all the other Accenture's branches. Accenture has been named

one of the world's most ethical companies for 10 consecutive years by Ethisphere (Ethisphere, 2019), and its code of ethics is an important part of this success. The Accenture Code of Business Ethics is a regulatory code of ethics that is build upon the core values of the organization. The six core values featured in the code (Accenture, 2019a), shape the diverse culture of Accenture. The values stated in the code are described next:



Client Value Creation: Accenture will enable their clients to become high-performance businesses by being responsive, relevant and constant in delivering value



One Global Network: Collaborate, make a healthy relationships and leveraging the power of global insights to deliver excellence.



Respect for the Individual: Accenture is committed with the creation of an open and inclusive

environment and treating each person in a manner that reflects the company's values.



Best People: The organization is focused on attracting and developing the best talent in a mutually supportive environment.



Integrity and Stewardship: The organization will inspire trust by being ethically undistortable, honest, and accountable. Moreover, in order to build a sustainable organization, Accenture commits to improve their employees lives and the community around them.

Accenture's code of ethics objective is to assist the company's employees to make ethical behavior a natural part of their everyday activities. It features also a new section that outlines the company's strategy to develop secure, transparent, and responsible AI systems. As stated by Chad Jerdee, General Counsel & Chief Compliance Officer of Accenture: "Our Code isn't just a document...it's what we believe, how we live and how we lead. It's totally embedded in all we do. It's how we improve our business performance and build on Accenture's reputation in the marketplace. It's how we put our clients and our people first. It's the Accenture Way of putting our integrity into action...every one of us, every day."

Additionally, Accenture developed a chatbot that provides easy and anonymous access to the ethical information of the code. The chatbot called COBE, was developed using Machine Learning models and it aims to help employees look at ethics in a new and more interactive way.

3.4.3 The economic value of Ethics

Ethics is projected to be one of the defining issues of the coming decade due to the fast pacing developments of AI (Étienne, 2019). One of the

main reasons of this, besides the alignment with the ethical practices of organizations, is the business value that ethics would bring in the future. Traditional business ethical approaches emphasize compliance with rules and systems. Many commercial institutions have focused on building a tightly controlled "culture of compliance" out of the concept of ethics, which is something that has scaled upon AI implementation and that could bring problems in the future. Furthermore, even if there is a huge interest from organizations to implement ethical practices as an added value proposition, sometimes the value of ethics is not well understood. For instance, there is a current consumer trust crisis that is affecting corporations like Facebook or Google due to the recent scandals around their unethical data policies (Francis, 2017; DeMers, 2018). Another example are the costs associated by "poor conduct" in the global banking sector, where the payments in fines for unethical behavior goes as high as £264 billion pounds. (Jill, 2017).



Figure 19: Accenture's digital Code of Business Ethics bot, better known as COBE (Fjord, n.d.)

The examples presented show that ethics has an actual tangible value, mostly to help companies prevent unintended and harmful outcomes that can affect their public image and, therefore, their earnings. It is a fact that trust, which is included into modern ethical best practices, increases profit margins (Blasingame, 2016; Schwab, 2014). Moreover,

according to Boddington (2017), ethics should be seen in a more positive way as its implementation would promote new opportunities for innovation. Therefore, in the long term, an ethical strategy towards AI would give a sustainable competitive advantage for responsible organizations.

3.4.4 Discussion

In this section, the strategic & commercial value of ethics was discussed. Unfortunately, the value of ethics in the business environment is often overlooked due to its complexity and its unjustifiable negative image. Nevertheless, companies like MOBGEN | Accenture Interactive actively promote a non-commercial commitment to moral values under codes of ethics and social responsibility campaigns. This shows the importance of ethics as a defining issue in this era and its actual strategic value. Ethics can help companies prevent unintended and harmful outcomes that can affect their public image, and also, generate an increment in their profit margins. Because of this, companies should consider ethics less like a “compliance” issue and more like an innovation driver.

3.5 CHAPTER CONCLUSIONS

This chapter spins around overview of Ethics in technology, as well as the ethical culture within MOBGEN | Accenture Interactive. Unfortunately, most of the theory behind the Ethics of Technology is currently too theoretical, however, several organizations have conceived codes of ethics in an attempt of making a tangible set of behavior regulations. Implementing a code of ethics alone is not enough as a more integral strategy should be taken by developing and implementing an ethical culture and an ethics program within organizations. In the case of MOBGEN | Accenture Interactive, the company follows a Code of Business Ethics that complements its ethical culture, which has been recognized as one of the best ethical efforts in the world. The value of ethics in the business environment is often overlooked due to its complexity. Nevertheless, it has been previously demonstrated that ethics can help companies prevent unintended and harmful outcomes at the same time be a drive for innovation, eliminating the “prescription” notion that is commonly associated with the term. A lot of important theoretical learnings were acquired in the literature review stage of the project as previously presented. Because of this, empirical studies were also performed in an effort to complement the knowledge gathered and it is described in the next chapter.

3.6 TAKEAWAYS FOR MOBGEN | ACCENTURE INTERACTIVE



MOBGEN | Accenture Interactive Employees **should embrace ethics as a discipline capable of generating commercial value** by driving innovation, not only with a “compliance” notion. For example, by developing ethical strategies to assess collaboratively the design briefs proposed by clients, even if there are not AI related.

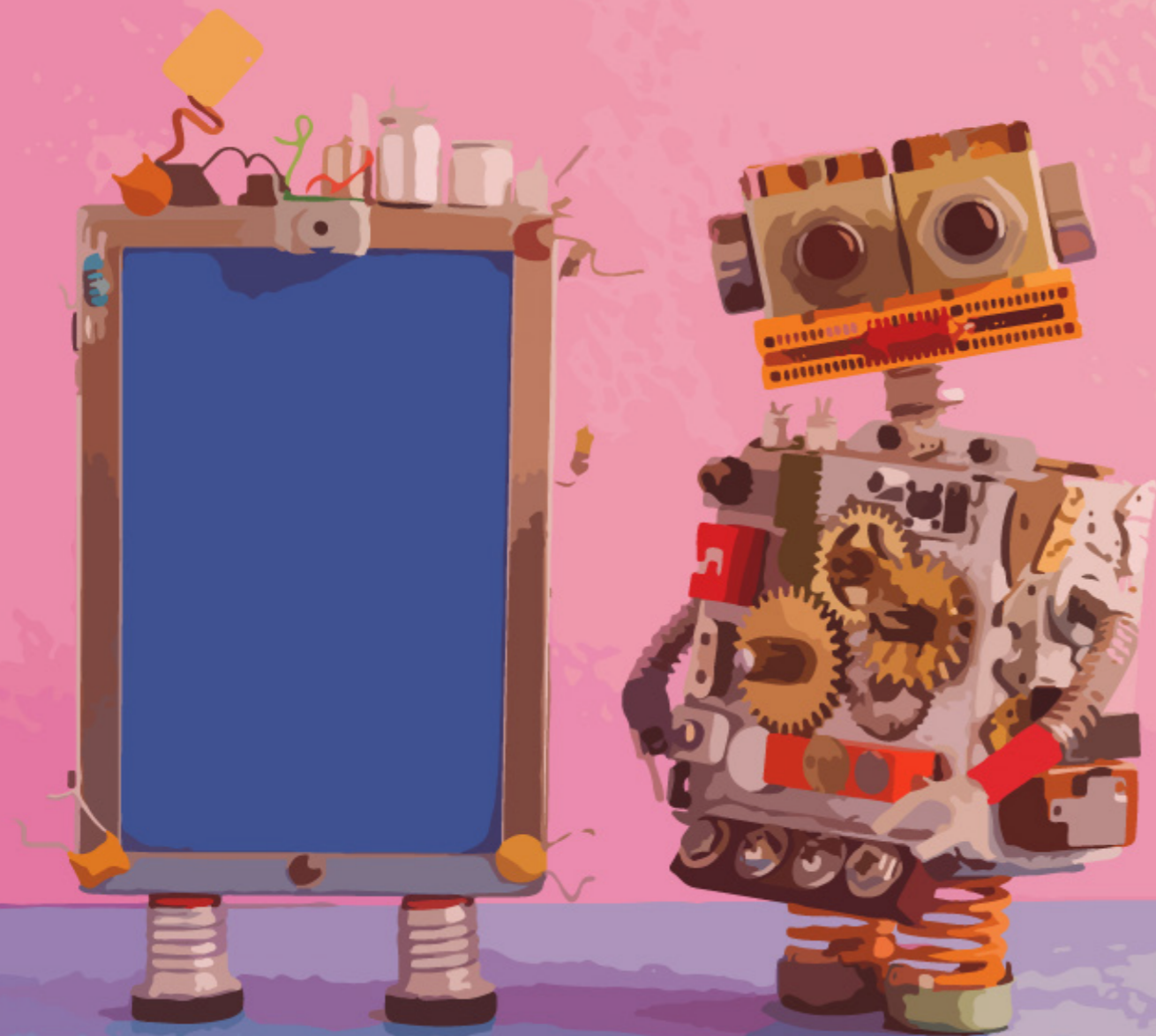


There are **three different normative ethics approaches** that could lead ethical decision making, which are deontological, consequentialist (or utilitarian), and virtue. These are better illustrated in Figure 14.



There are **several guidelines that touch upon the ethics of AI**, which are continuously improving as the development of AI continues to thrive. It is advised that employees from the organization to check them continuously in order to gather more knowledge on the topic. Some examples include the “Everyday Ethics for AI” guide by IBM and the People+Human Guidelines by Google.

04



chapter 4

Design Research for Context Exploration

The current project features a double diamond design process that includes an action research cycle on each diamond. Several exploration methods were used to gather insights for the understanding of the current context of the ethics of AI and are discussed in this chapter. The methods used include semi-structured interviews, generative tools, customer journeys, and stakeholder analysis, among others. Also, during this stage, an action research cycle was performed in order to learn more about and transform the popular perception of the ethics of AI.

4.1 RESEARCH SETUP

This section describes the research approaches that were applied in this thesis to explore the existing context around the ethics of AI. The next sections elaborate on the research methods separately, analyzing the goals, structure and the resulting insights gathered.

4.1.1 Empirical research Setup

It is important to explore the current state of the development of AI applications and the ethics of AI before ideating a solution. From the literature review it was concluded that a responsible approach to the development of AI applications is needed. Furthermore, as the development road of an AI application involves a larger number of stakeholders, it is expected that a more integrative research procedure would have better results. Because of this, and in order to complement the systematic literature review and the prior analysis of the ethical context around AI, an empirical research stage was performed. This stage consisted on several design research methods made specifically in order to answer the first two research questions.

This empirical research began with a series of 14 semi-structured interviews with ethics of AI-

experienced stakeholders, from AI users to Ethics of Technology experts. The main objective of this was to explore the current ethical view of the stakeholders involved in the development of an AI application. After this, a creative discussion with AI researchers from the TU Delft was performed in order to understand the ethical concerns that exist specifically with this type of stakeholders, as they are the ones that actually generate the state-of-the-art models that might be included in a commercial application in the future. This phase ended with the presentation of a provotype on an “Ethics & Design” seminar where the main objective was to communicate and to create awareness regarding the failures of AI commercial applications. Most of the interviewees that participated during the empirical research stage were consulted along the whole project as experts on the topic.

Some of the most important insights found in this stage highlight the lack of a proper understanding of ethics from most of the stakeholders involved, although, it was noted that there is a common interest on pursuing a responsible approach towards the topic. Certainly, it was also noted that MOBGEN | Accenture Interactive is currently unprepared regarding the ethics of AI and the possible unintended consequences that an unethical implementation of AI could bring.

4.2 SEMI-STRUCTURED INTERVIEWS

A series of 14 semi-structured interviews was conducted to explore the current ethical view of the stakeholders defined in previous chapters.

4.2.1 Interview goals

The participants have a broad range of occupations and professional backgrounds, being all of them relevant stakeholders within the context of the Ethics of AI. Participants occupations included: Ethics in Technology and AI Researchers from the TU Delft, AI Implementors from MOBGEN | Accenture Interactive, AI enthusiasts, AI entrepreneurs (Envision startup), and AI-products users.

The main goals for this stage comprised:

- **Understanding participant's opinion regarding AI.**
- **Defining the actual value that AI applications have in their lives.**
- **Discovering the participant's personal perception of ethics.**
- **Defining the ethical considerations and concerns participants have about AI.**
- **Understanding participant's perception of failure as a mean for unintended ethical consequences (consequentialist approach).**

4.2.2 Research approach

Prior research shows that due to its versatility, semi-structured interviewing is one of the most popular methods of data collection and knowledge production in academia (Kvale & Brinkmann, 2009). Moreover, this method is often used in software development research and new product development projects (Hove & Anda, 2005; Sandmeier et al., 2004). Therefore, it is an adequate method for the contextual study of AI ethics. The interviews followed the recommended general guidelines for semi-structured interviews proposed by Patton (2002), and an interview guide was prepared and modified accordingly during the study. The interview guide followed is shown in Appendix A.

The interviews started with more common topics such as job description and their actual experience with AI. All fourteen interviews were audio recorded and notes were taken for further analysis. All the interviews were transcribed, sorted, and classified in statement cards (Visser, Stappers, Van der Lugt, & Sanders, 2005; Sanders & Stappers, 2013). These cards were clustered to find patterns and collect

insights in order to answer the research questions and verify the assumptions made at the early stages of the project. A more detailed overview of the cards is shown in Appendix C.

4.2.3 Gathered Insights

The performed interviews allowed to answer the first research question of the project. Moreover, it concentrated the ethical views of all the stakeholders involved, as well as their expectations and concerns towards AI and its future. Some of the most important insights are addressed next:

- **A lack of a proper understanding of ethics was discovered. Most of the interviewees were able to define ethics and show their ethical views, however, they also tended to confuse ethics with morals as using a definition around ethics of "what's right or wrong". This definition limits the ethical spectrum towards a very narrow moral view.**

- **It was noticed that most of the interviewees that have an active role in the development of AI, applied personal ethical understanding and principles to the development of the AI projects they were involved in. Something that has been previously addressed in research.**

"Most of the time I do look for biases and privacy issues on the datasets I work with...we don't have a special procedure for that, it is something I do because it is the right thing to do..."

Software Developer - MOBGEN | Accenture Interactive

- **A general overview of the ethical concerns regarding AI was addressed. It shows that most of the interviewees are aware of the ethical considerations that as humans we must enhance or take into account while designing AI systems.**

"I think ethics is basically do something in the best interest of view and group around you or your social environment, something like that"

Android Development Manager - MOBGEN | Accenture Interactive

"Ethics it's like acceptable behavior by society".

AI Expert/Creative Technologist - MOBGEN | Accenture Interactive

"Ethics, I have one way to measure whether I'm good or not. And that's my conscious. Okay, if I feel bad afterwards, after doing something that is against my ethics, if I don't feel bad about it, then it's still Okay."

AI user / Data Management and Visualisation Consultant - MOBGEN | Accenture Interactive



Figure 20: Semi-structured interview with an employee from MOBGEN | Accenture Interactive (picture by author, 2019)

“Well, you need to you need to make sure that the AI not gonna rise like ethical discussions like you may need to make sure that your AI is not racist or discriminates minorities, or because that’s not ethically approved these days. So that’s the things that you need to have in consideration.”

AI Expert/Creative Technologist - MOBGEN | Accenture Interactive

“...nowadays is like people assume that AI is somehow this unbiased thing. But the problem is the data that you’re feeding, it affects how biased on bias it is, I mean, essentially, literally, they are teaching it biases, right to filter out the noise (...) And think thinking that it’s a tool that is completely unbiased when in reality is really not”

Creative Technologist - MOBGEN | Accenture Interactive

- Some interviewees showed a poor understanding of the social consequences of AI, for example some AI implementors mentioned cases related to the “relativity” of ethics . On the other hand, some AI users did not feel concerned regarding privacy issues.
- There is no current ethical assessment for any of the design briefs that deal with AI within MOBGEN | Accenture Interactive. The main reason for this is the lack of awareness of the possible consequences and the lack of tools to make the assessment within the company own ethical standards. Moreover, some employees think that customers do not have enough knowledge and maturity on the field of AI and ethics to convey on a requirement different than GDPR compliance.

“Currently we have no way to ethically assess a brief...I mean, the only discussion around ethics we have is when the team feels that something ‘is off’ with the project.”

Innovation Lead - MOBGEN | Accenture Interactive



Figure 21: Semi-structured interview with an employee from MOBGEN | Accenture Interactive (picture by author, 2019).

- The collected insights supported the development journey of an AI application stated at the beginning of the project. A more detailed process was also defined with the validation of several experts from all the stakeholder groups involved in the research stage, including AI implementors/experts from MOBGEN | Accenture Interactive, AI enthusiasts and AI researchers.



Figure 22: A sample of the statement cards used for this research project (picture by author, 2019).

- One of the most important results of this stage was the characterization and re-mapping of the stakeholders involved in the development of AI applications. This was made by a synthesis of the insights from the interviewees that were translated into three different personas, AI researchers, AI implementors/clients, and AI users. A more detailed view of the personas is shown on Appendix D. The personification of stakeholders help with the understanding of the general ethical views and needs and goals of the stakeholders involved in the

“Like ethics is such a broad question, right? I mean, when you’re dealing with scientists and engineers, they don’t like such vague notions of like morality, because it’s not something you can hard code is something that’s subjective”

Creative Technologist - MOBGEN | Accenture Interactive

If you talk about safety concerns, it’s my own opinion but I don’t care about privacy, because I believe there is no such thing anymore.

AI user / Data Management and Visualisation Consultant - MOBGEN | Accenture Interactive

development process of an AI application. Furthermore, this was used to add an extra layer to the theoretical framework proposed on Chapter 5.



Figure 23: A sample of the first clusters generated with the statement cards around the ethics of AI (picture by author, 2019).

4.2.4 Discussion

The empirical research shows that, although there is a considerable awareness of the ethical considerations regarding the implementation of AI, it seems that ethics is still treated as a superficial topic. Most of the interviewees differed about an extensive definition of ethics which demonstrates the complexity of the topic and lack of enough education around ethics. Furthermore, it was found that some of the interviewees experience a feeling of “passive responsibility” (Van de Poel & Royakkers, 2011), which is something that should change in order to have an ethically and socially desired AI development. An important insight to consider is the absence of current design tools to assess the ethics of a project within MOBGEN | Accenture Interactive. One of the main causes of this appears to be that, similarly to design, ethics does not bring a tangible value to an organization, so it is often overlooked. Considering that this type of insights might be reflected on other stakeholders as well, a creative discussion was organized with the participation of AI researchers from the TU Delft.

4.3 CREATIVE DISCUSSION

A creative discussion regarding the ethical context in AI research was conducted with the support from six AI researchers of the EEMCS Faculty of the TU Delft in order to understand the ethical concerns that exist specifically with this type of stakeholders. The discussion was organized due to the current interest that a group of researchers from the EEMCS Faculty of the TU Delft shown towards the topic of this research project.

4.3.1 Discussion goals & approach

This discussion took place in building 28 of the TU Delft, which is part of the EEMCS Faculty. Insights of the recordings were transcribed in statement cards (Visser,et al., 2005; Sanders & Stappers, 2013). These cards were added to the interviews insights in order to convey in a more robust analysis applying a true multi-stakeholder approach. Furthermore, a brain writing exercise was performed to get a general understanding on the definition of ethics that the researchers have.

The main goals for this stage comprised:

- Understanding the researcher's opinion regarding the ethics of AI.
- Discussing the process followed by the researchers to come up with a model.
- Identify if ethics is considered and what type of ethical flags are raised.

4.3.2 Gathered Insights

The creative discussion helped to define the role of AI researchers in the ethics of AI, as well as the process followed by academic research in the field. Some other insights are briefly addressed next:

- A lack of understanding of the ethical consequences that AI could have was discovered. Although the discussion was supported by a group of researchers deeply interested in the topic of Ethics of AI, it was noticeable that there was a limited idea of the consequences of their research could provoke, as well as a misaligned definition of ethics. Thus, it could be stated that there is a need for a design tool that copes with this type of ethical unawareness.



Figure 24: Result of brainwriting session (picture by author, 2019)

"Now, there are some ethical stuff that we kind all agree upon right? But I don't think that is helpful in this kind of situations."

Associate Professor - EEMCS Faculty TU Delft

"I think we are overthinking this thing about ethics too much, don't you think? It is getting too philosophical..."

PhD student - EEMCS Faculty TU Delft

"I think that ethics is related to the outcome itself, the report that is taken as facts. Not the research question we propose, otherwise why do we even make this research?"

PhD student - EEMCS Faculty TU Delft

"How close is what you do from humans? I think that it is also one of the most important things, the closer your research to actual human data, you need to be more ethical"

Postdoctoral Researcher - EEMCS Faculty TU Delft

- Most of the researchers gathered were aware of the GDPR considerations during their research experiments, nevertheless, several ethical flags, were discovered while discussing the research process followed. This made the researchers discuss and reflect upon their own participation within the field as to think about the question "Is our Research Question ethical enough to start with?". This strengthens the need of a design solution that also teaches what could be considered ethical depending on the context of the stakeholders.
- The lack of integration of other stakeholders for the ethical considerations and for further implementation of the results was addressed also. Tests within human environments follow the standard ethical procedures authorized by the TU Delft, nevertheless, there is no inclusion of the participants in the definition or reflection of the ethics of the research, ethics is represented to the participants only in a consent form. Moreover, the ethics of the research lays on the researcher personal considerations, as long as the GDPR regulations are followed. Likewise, after the research effort is done and published, the research group feels responsible for future usage of the results, although not every researcher thinks the same. This, as mentioned by the team, is "not ethical, because people tend to treat this

results as facts, but there are several research limitations that people are not aware of...". Hence, it is crucial that AI researchers take into consideration most of the stakeholders impacted, initially or in the future, by their results at all times during their research.

"Yes, we deal with GDPR compliances as kind of an 'ethical assessment' for our research, as we mentioned before, it is the standard... we currently don't have any other ethical discussion around during our experiments."

Postdoctoral Researcher - EEMCS Faculty TU Delft

"Yeah, but what they do with our research is not ethical is currently out of our reach... we need funding to keep working, don't forget about that"

Postdoctoral Researcher - EEMCS Faculty TU Delft

"My research don't deal with people, so I am not interested on this"

Head of the Computer Vision Lab - EEMCS Faculty TU Delft



Figure 25: The creative discussion took place in "Building 28", located in TU Delft main campus.

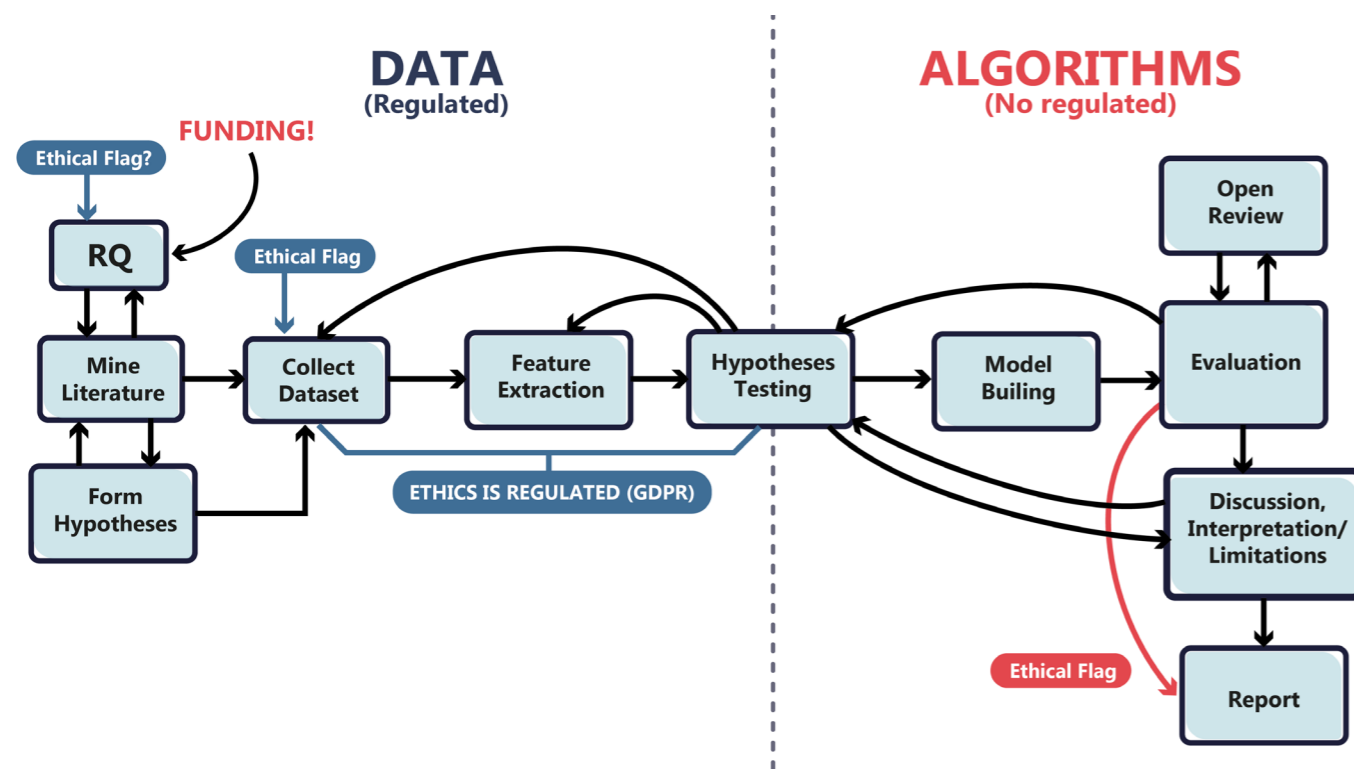


Figure 26: Visual adaptation of the process built with the AI researchers (illustration by author, 2019)

4.3.3 Discussion

The world of academia is not that distant from the professional environment and this section shows why. Certainly, some patterns repeat as the feeling of passive responsibility, or not responsibility at all, is also experienced by AI researchers. Moreover, the non-inclusion of other stakeholders (i.e. users), on their process aligns with the current notion of ethics as a “prescriptive” discipline. Nevertheless, this study also shows that the AI researchers of the TU Delft do follow ethical legal requirements protocols like GDPR forms. So, it is expected that by providing adequate tools for ethical understanding, this researchers might include an ethical “double check” during their process to get to a more “active responsibility”. Furthermore, at this point of the empirical research a considerable large amount of knowledge have been collected. This matched with the organization of an “Ethics & Design” seminar by the Honours Programme community of the TU Delft. Taking advantage of this opportunity, a provotype was exposed as part of a research progress results presentation.

4.4 HONOURS PROGRAMME SEMINAR: PROVOTYPE DISCUSSION

This section states the presentation of a provotype on an “Ethics & Design” seminar with the objective of communicate and to create awareness regarding the consequences of common failures of AI commercial applications.

4.4.1 What would you do if...?

To complement the research stage, a “provotype” discussion session was conducted during an “Ethics & Design” Seminar organized by the Honours Programme community of the IDE Faculty of the TU Delft. A “provotype” is a type of prototype that exposes tensions of a field of interest in order to trigger participatory analysis and collaborative design innovation (Boer & Donovan, 2012). A provocative scenario was presented to the attending audience of the seminar, as shown in Figure 27, in order to expose ethical tensions related to an AI-powered product (Google Home). The presentation was followed by an enriching discussion where written and spoken feedback was collected from the crowd. The audience was conformed mainly by TU Delft students and IDE Faculty professors.

4.4.2 Provotype Goals

This provotype was presented with a clear goal in mind, to collect feedback from the users about the tension the scenario provokes. By sparking the imagination of the audience with a provocative statement, they would be able to question the ethics of the designers that work on the UX of smart assistants. The discussion followed the approach proposed in Chapter 1, where a consequentialist approach was proposed with the intention of derive ethical principles from AI-failures (consequences of unethical AI).

4.4.3 Gathered Insights

The creative discussion triggered interesting comments from the attending audience. Some of the most important insights were:

- The ethical issue was considered by a considerable part of the audience as a

context design failure from the designers and developers of the smart assistant. This strengthens the idea of considering ethical consequences during the design and development of AI applications.

- Some of the participants of the discussion argued that the ethical failures were the responsibility of the developing companies, but some other mentioned that the user plays also an important role regarding the perception of the failure and its ethical consequences.



Figure 27: Presentation of the ‘provotype’ during the “Ethics & Design” seminar (picture by author, 2019).



Figure 28: The 'provotype' presented during the "Ethics & Design" seminar (illustration by author, 2019)

4.4.4 Discussion

The topic of ethics is usually presented by creating a controversial and provocative statement which is intended to make people question their own ethical decision making. In the case of this "provotype", the usage of an unforeseen interactive failure generated a discussion where attendees discussed their ethical views and how this failure would have been taken by them. This discussion gathered interesting insights as the field of context failures in AI is currently unexplored but plays an important role in the ethics of AI.

4.5 ACTION RESEARCH CYCLE 1 - LEARN

As commented previously, an action research cycle was executed simultaneously to the empirical research, which is explained in this section. The cycle features four different phases: plan, act, observe, and reflect (Price et al., 2018). It is important to mention that some of the actions performed are also part of the empirical research efforts.

4.5.1 Plan & Act

The first action research cycle applied in the project was planned with the objective of understanding the current context of the ethics of AI, as well as to try to transform the common perception around it. This was achieved by performing a certain number of actions during the Discover and Define phases of the project. This chapter covers some the actions followed on the former, which include the introduction of a creative discussion executed with AI researchers, a presentation done during an "Ethics & Design" seminar, and several informal talks with experts and enthusiasts on the topic (some made after the semi-structured interviews took place). An relevant action worth to mention was the invitation that the author of this thesis got from a group of Accenture interns to help with the organization of an internal event about "Human-AI". This event gave the opportunity for the author to spread the view on ethics around other branches of the organization as content manager of the "Responsible AI" that was covered during the event.

4.5.2 Observe & Reflect

It was observed that the actions performed during this stage had an interesting effect on people. For instance, the interviewees that work at MOBGEN | Accenture Interactive have a deep interest on the topic of the ethics of AI, but do not count with any type of tool or method for ethical assessment for their projects. Moreover, phrases like "what is ethical and what is not is quite subjective" were mentioned several times during the lapse of the empirical research either by some interviewees or by informal chats with other Accenture employees from diverse departments. This confirms the insights gathered from the empirical research

efforts and literature review, where it was noted that ethics is a topic that is often taken as superficial and "relativistic". Nevertheless, according to the information gathered with experts, this attitude is one of the biggest dangers that we face towards the development new technologies as we need to enhance an culture of ethics through education.

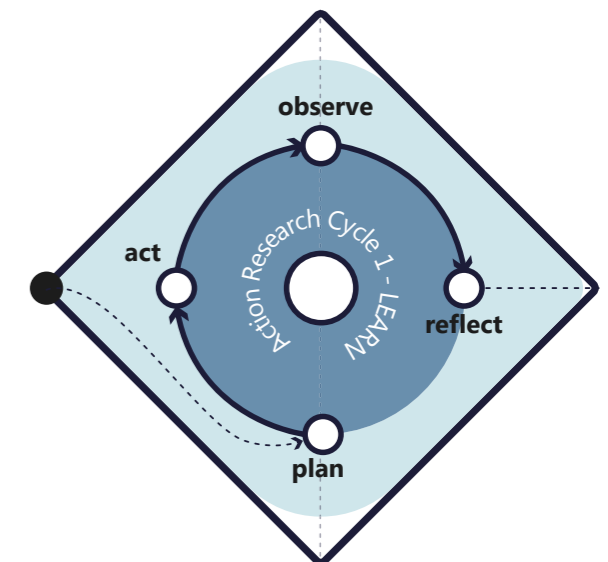


Figure 29: Firts Action Research Cycle as presented in Chapter 1 (illustration by author, 2019).

"The educational system is failing engineers to train them in understanding that what they're doing is not well you know... this is the same as the debate about that guns don't kill people, people kill people. That is total nonsense. You as an engineer are responsible OK, period. Now what is the level of your responsibility is a debate we're not going to get into that right now(...) but then I go back to the education part. Education is really, really fundamental."

Postdoctoral Researcher - TPM Faculty TU Delft

The results observed from this stage worked as a motivation to include an “educational” sense to the tools designed during the project. It is clear that people is interested in the ethics of AI and would like to implement ethical practices around its development. However, the lack of proper understanding and education on ethics and the lack of tools to spark an ethical conversation and propose solutions within the AI environment makes it difficult to contribute. These results were used as a base for the creation of a theoretical framework for the Ethics of AI, which is futher developed in Chapter 5.

4.7 CHAPTER CONCLUSIONS

This chapter describes the research methods used and the insights gathered during the empirical research stage. It also describes the process followed during the first action research cycle of the project. The empirical research shows ethics is still perceived as a superficial topic. However, there is a noticeable interest and awareness from most of the interviewed stakeholders on the ethical considerations of implementing AI. Moreover, a feeling of “passive responsibility” is present in stakeholders like AI researchers and implementors. A majority of the insights gathered align with the current notion of ethics as a “prescriptive” discipline, which is something that needs to change according to experts in the topic. An important finding of this stage is the absence of current design tools to assess the ethics of AI related projects. Certainly, this results demonstrate the existence of a window of opportunity for the development of ethical tools that assist in the responsible implementation of AI and the current business value that this might bring in the future. With all the gathered insights, the next steps to follow is to establish a theoretical ground that could be used to generate ethical solutions for the development of AI applications. This has be done with the generation of an Ethical Framework for AI, as it is discussed on the next chapter.

4.8 TAKEAWAYS FOR MOBGEN | ACCENTURE INTERACTIVE



There is a **considerable interest in and awareness on the ethical considerations of AI** applications, however, ethics is still treated as a superficial and ‘relativist’ topic. This shows that a **lack of education on societal and business ethical consequences is present** in MOBGEN | Accenture Interactive employees.



Most of the stakeholders that have an active role in the development of AI, **apply personal ethical understanding and principles to the development of their AI projects**. This is something that is not desirable as does not includes an ethical discussion with other parties to get a common ethical consensus.

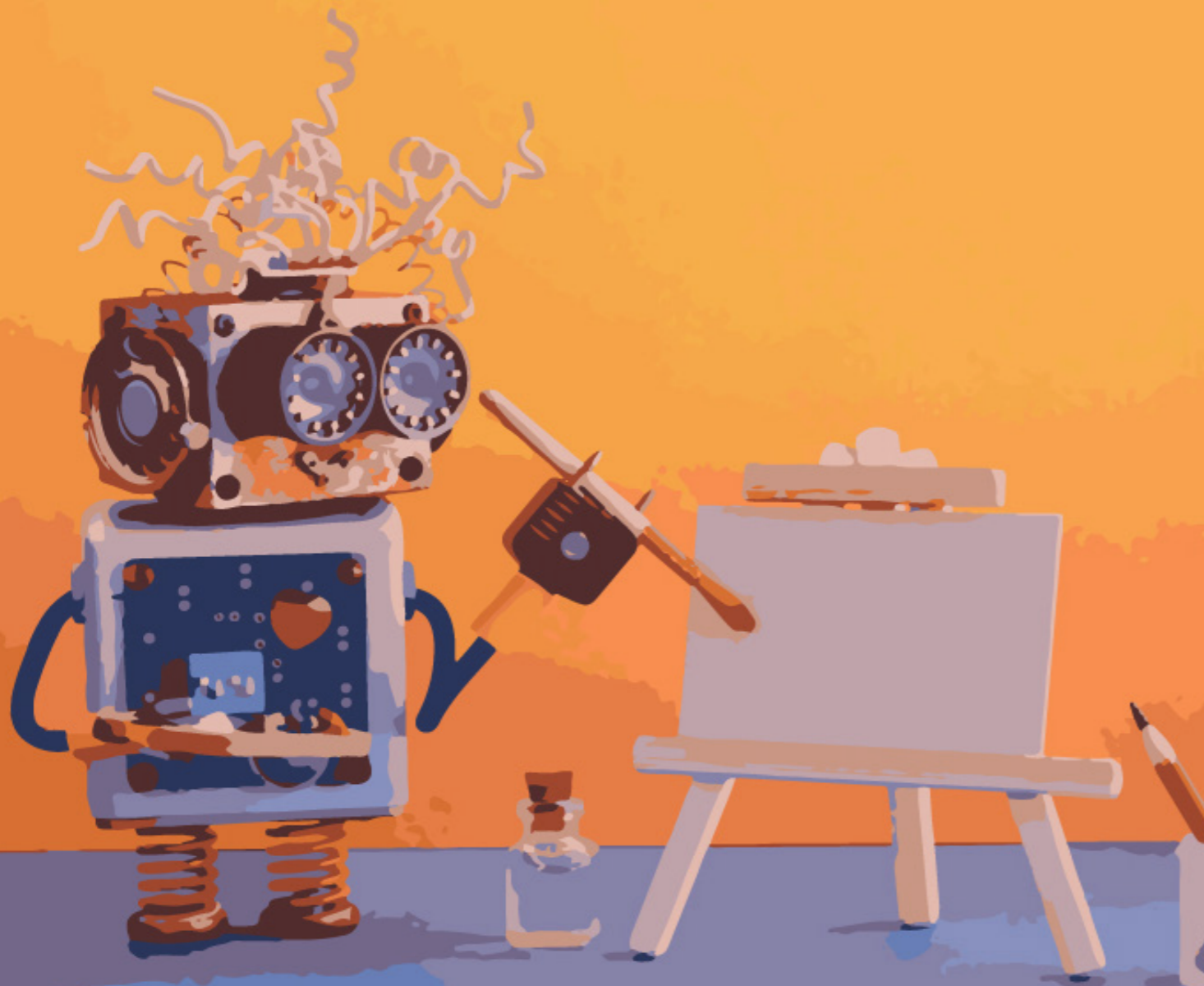


It is common for AI developers to have a **feeling of “passive responsibility”**, which is not desirable as it only provide accountability if something fails. **An active responsibility attitude should be taken by all the employees of the organization** that work on AI-related projects as a shared responsibility exists on this implementation by all the participating stakeholders of the process (AI researchers, implementors, clients, users).



An important insight to consider is that **there are currently no tools to assess the ethics of a project within MOBGEN | Accenture Interactive**. Hence, a tool for doing so would be a good improvement to the organization’s value proposition.

05



chapter 5

Transforming Insights Into Design

This chapter describes the development of an Ethical Framework of AI based on the results of the literature review and the empirical research stage. First, a more concise version of a development journey of AI applications is discussed, then a stakeholder analysis process and the making of personas out of the different stakeholders are analyzed. Furthermore, by applying these and other design methods, an ethical framework was conceived in order to introduce and support a solid theoretical ground to assess the ethics of AI applications. The framework was embodied using a “puzzle” metaphor and was also validated afterwards.

5.1 DEVELOPMENT JOURNEY OF AI APPS

As a result of the collection of insights on previous stages, a more detailed overview of the development process followed for an AI application is presented. This section discusses this development journey and how the insights collected assisted for the creation of a theoretical framework.

5.1.1. The journey of an AI Application

One of the results of the interviewing stage was the definition and validation of the development journey that an AI application follows. This was proposed at the beginning of the project as a result of the literature review and some assumptions made. A simplified version of the process is shown in Figure 30. The process generally starts with the AI researchers as they are the ones that create the state-of-the-art AI models and algorithms. These researchers could work on academic environments or in commercial ventures like Facebook or Google (Matney, 2018). It is important to mention that AI researchers need a considerable amount of data to test their work, therefore, this data collection process raises the first ethical flags of the journey.

Along the whole development cycle, there are other ethical checkpoints derived from the GDPR compliance policy applicable only in the EU. This is the specific case of the AI implementors, where only GDPR compliance checks are specified at the beginning of the project and confirmed before scaling up the application. This leaves the assessment of the rest of the ethical considerations to the developers' and designers' own criteria. This is the approach that reigns within MOBGEN | Accenture Interactive projects. In the end, the AI application is delivered to the final user who could be impacted by unattended ethical flags along the journey.

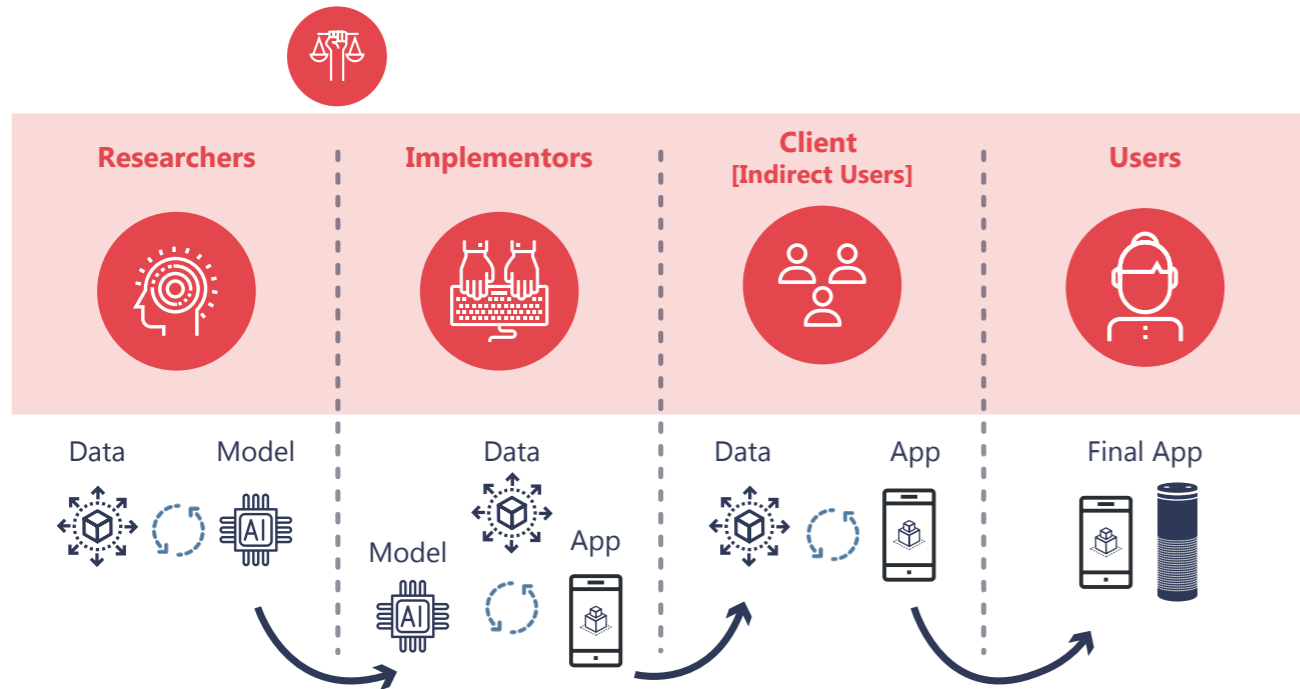


Figure 30: A more refined visualization of the Development Journey of an AI application (illustration by author, 2019).

5.1.2. "One stop" AI developers

Some organizations englobe the whole spectrum of development. These companies count with a group of in-house AI researchers and they act as implementers and clients for a B2C audience, for example, social media platforms like Facebook and digital giants like Amazon, Alibaba, and Google. On the other hand, due to its B2B model some companies only work on the first two stages of the development process, like IBM or Microsoft. Organizations like Accenture & Mckinsey & Co. also fall into this category as consultancies mostly follow a B2B model.

5.1.3. Ethicists in the AI Journey

As mentioned by various of the interviewees, ethics is present in the development journey only on a personal basis by the developers and designers on each stage. This means that in general, each stakeholder is responsible of the ethics of their work, which makes the ethical situation around AI more complex. Moreover, the only considerations taken (if so) are those related to GDPR. Most of the experts of Ethics of Technology interviewed agree on the idea of companies involving ethicists within their structure, as they find essential to have an in-house ethical assessment for a responsible AI development. Nowadays, some organizations have started to incorporate experts in ethics into their business structure as it is the case of Mckinsey & Co. (Mckinsey, 2019).

4.4.4 Discussion

The journey of an AI application proposed at the beginning of the project was not that far away from reality. This journey begins with the development of the state-of-the-art AI algorithms by AI researchers and continues towards the future implementations of these on commercial applications. It is important to mention that until now, ethical efforts have only been proposed on the process of the creation of these algorithms. However, it has been proven that this effort is not enough and more "ethical checks" could be performed during the implementation stage, which is the promise on what this thesis is built upon.

5.2 STAKEHOLDERS MAPPING

As commented previously, an action research cycle was executed simultaneously to the empirical research, which is explained in this section. The cycle features four different phases: plan, act, observe, and reflect. It is important to mention that some of the actions performed are also part of the empirical research efforts.

5.2.1 Stakeholders overview of AI-apps

A redefined version of the stakeholder map of an AI app development resulted from the research stage which could be observed on Figure 31. It is important to mention that an addition was made to the relevant stakeholders with the inclusion of the government. The reason for this was the tight relationship that exists between ethics and policy making, especially on new technological developments, which is the case of GDPR.

It can be noticed that there is a complete disconnection between the AI researchers and users. This was addressed during the creative discussion made on this stage as the researchers currently do not have (or have a limited) contact with actual users. This brings up another problem related to the accountability that researchers feel while developing AI, which is something that also is reflected in the implementors insights. It is also observed that the role of the researchers in ethics of technology is pretty limited as they are currently not involved enough in the industry.

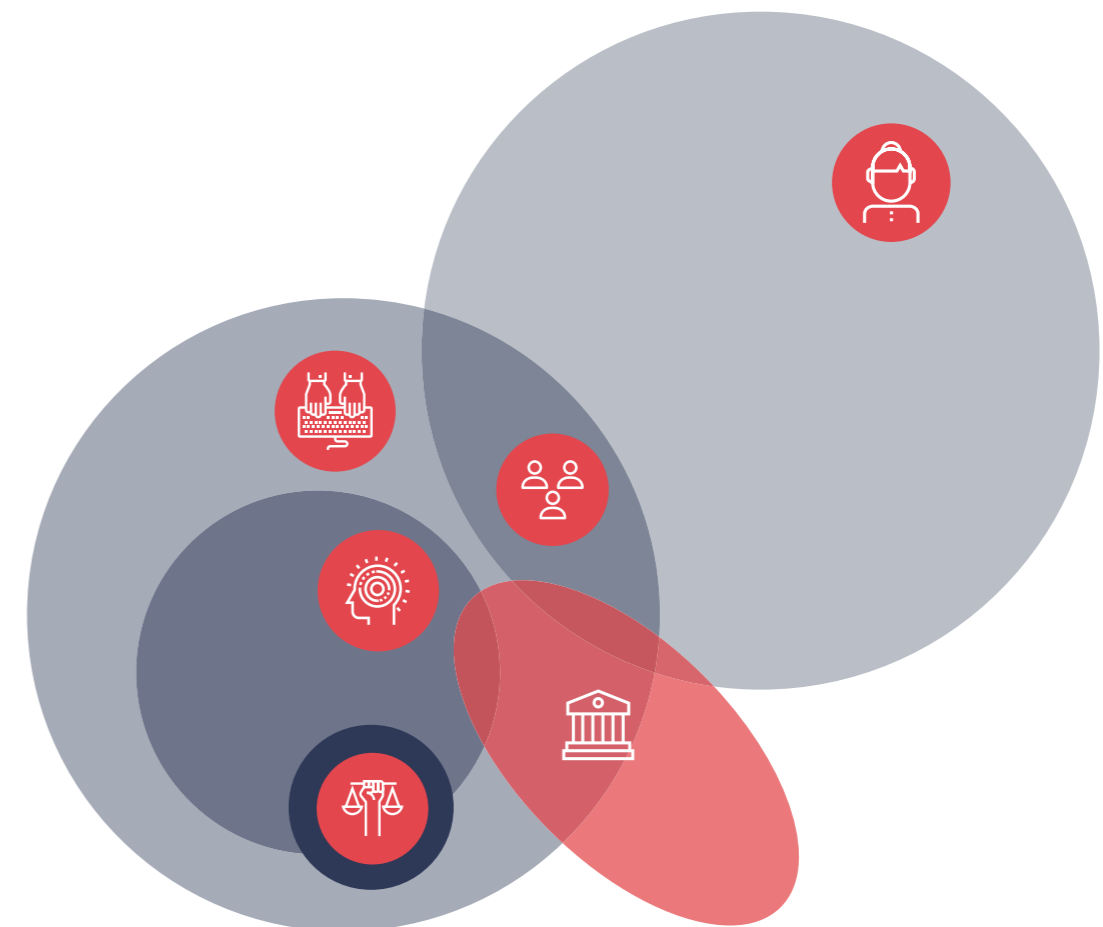


Figure 31: A more refined representation of the initial stakeholders map, where the role of the government & regulations is included (illustration by author, 2019).

5.2.2 “Personification” of Stakeholders

Several personas were made with the information collected from the interviewing stage in order to analyse and get a visual representation of the characteristics, ethical views, and needs of the main stakeholders. The “personification” of the stakeholders prompt the addition of the stakeholders to the proposed ethical framework for AI. This was useful to understand the general ethical views and needs and goals of the stakeholders involved in the development process of an AI application.

Margaret

AI Researcher
Post-doc @ TU Delft



Personal Characteristics

- Curious
- Strong technical skills
- Medium interest on the ethics of AI
- Stubborn and egocentric (“professor mentality”)

Drawbacks

- Unfamiliar with ethical concepts
- Takes ethics into account only during data collection.
- Needs to be “ethically flexible” with research to get funding for future research.

Needs & Goals

- Looks for recognition in the field.
- Aims to be ethical in her research due to own principles.

Ethical Views

- Accountability is important at personal level but not at research level.
- Ethical flags and interest only on:
 - Data privacy
 - Bias in data

Figure 32: Example of persona created based on the insights collected from the AI Researchers

5.2.3 Discussion

The addition of the government on the stakeholder map shows how connected the development of AI is. Furthermore, it also shows the importance of a legal framework for the implementation of ethics in the field of AI applications. Previous research tells us that a regulating body is important for an ethical development of technology, which is why the inclusion of the government creates a more robust stakeholders map. Moreover, this “deontological” approach to the ethics of AI would be used for the generation of a theoretical framework that is discussed more thoroughly in the next section

5.3 ETHICS OF AI FRAMEWORK

This section describes the steps followed for the creation of a theoretical framework for the ethics of AI. In order to conceive an ethical framework, several ethical principles were defined first using the insights collected from the literature review and empirical research stages. For this to happen, a heuristic approach was taken and several design methods were applied, for example, brainwriting and clustering. By applying these design techniques an ethical framework was set, which is expected to be used as a reflection tool for setting standards of conduct and behavior. The framework portrays an “ethical cycle” that could be taken into consideration during the development of AI applications. The framework would be used as a basis for the design of the final ethical toolkit.

5.3.1 From Insights to Ethical Principles for AI

In order to generate ethical principles out of the insights gathered, a series of brainstorming exercises were executed with the help of some MOBGEN | Accenture Interactive employees and students from the TU Delft. The ethical principles defined in this stage are referred to the ethical considerations that should be taken when developing AI-systems, for example, the accountability of an algorithmic failure. During one of the brainstorming exercises, clusters were formed using the collected ethical principles found in the literature and the ones recalled as vital by the participants of the empirical studies. The clustering helped to generate 9 main

ethical principles that could be applied to AI. The first principles were expanded to 11 after some discussions and validations with students from the TU Delft and Ethics in Technology experts. The AI ethical principles generated were defined using questions in order to also trigger ethical discussions later in the project.

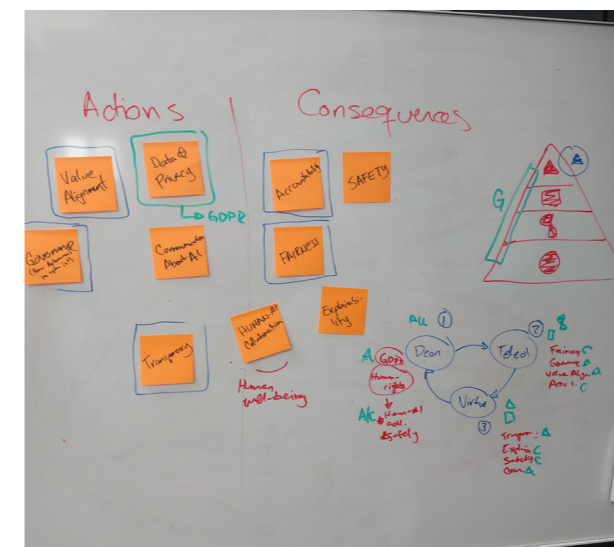


Figure 33: Sample of analysis made by classifying the ethical principles uncovered into actions and consequences in order to shape a theoretical framework.



Figure 34: Clustering of AI ethical principles defined during this stage.

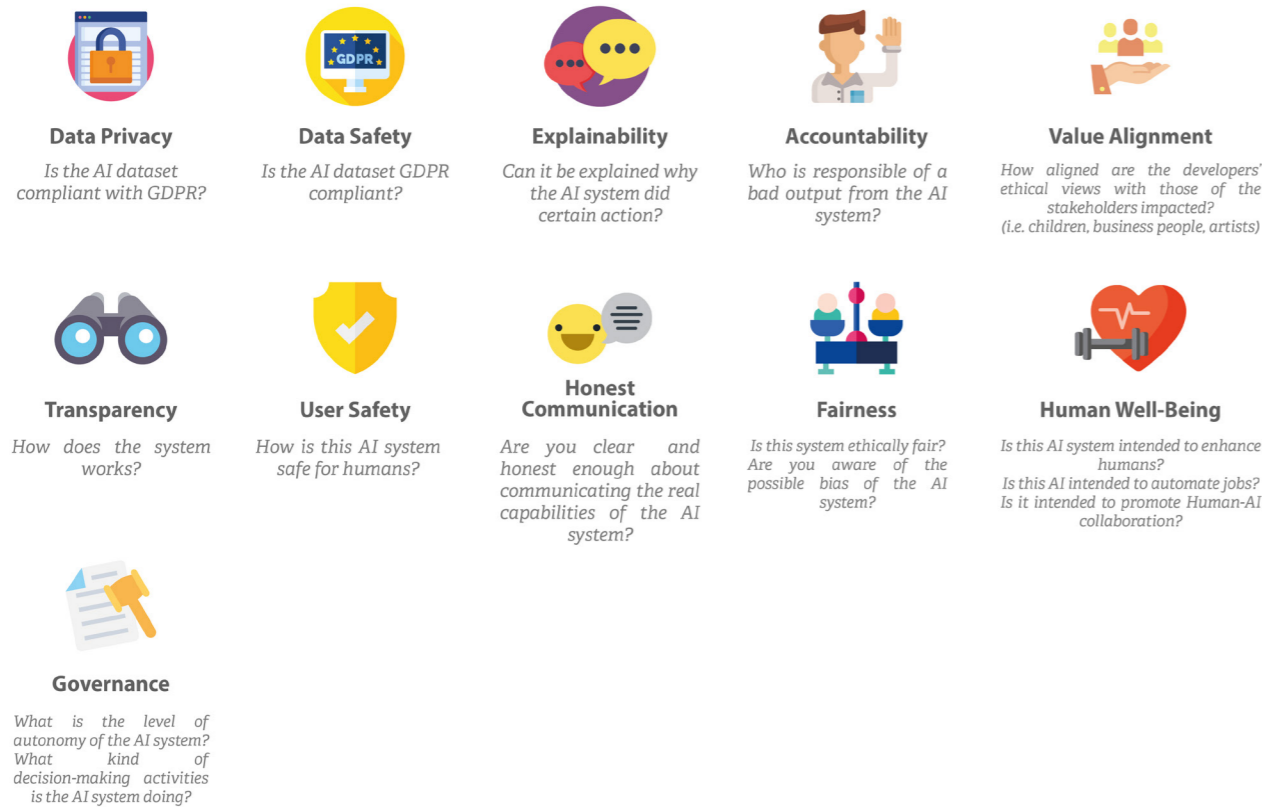


Figure 35: An overview of the definitions of the ethical principles defined with questions in this stage.

5.3.2 Building an ethical framework from an IoT approach

Once the ethical principles were defined, a theoretical framework that was still required. According to the literature, an ethical framework sets recognised standards of conduct and behaviour (APM, n.d.; Seneges et al., 2017; Bond & Firenze, 2013), for example, a consequentialist ethical framework defines what is ethically correct for a determined social group as a whole.

The literature review revealed that most of the current ethical frameworks for AI only cover certain epistemological terms in a non-visual way (Cutler, Pribic & Humphrey, 2018; Lubbock & Virdee, 2018; European Commission, 2019; Google, 2019). Moreover, most of the intended AI frameworks do not practically support decision makers as most of the guidelines published related to the topic convey only in reflective questions and not in an

ethical process to follow. Furthermore, companies like Mckinsey & Co. (2019) have expressed the importance of AI ethical frameworks that feature ethical processes. Because of this, several analysis sessions were made in order to come up with an ethical framework for the project, which are described in detail on Appendix B.

After the analysis stage, it was found that a similar approach towards ethics of technology has been developed previously by Seneges et al. (2017). As shown in Figure 36, the framework proposes the use of three different ethical approaches (deontological, consequentialist, virtue) to discuss the assessment of the ethics of IoT applications. With this composite framework approach, a new ethical framework was proposed in order to discuss the ethical dimensions of AI development. This framework will serve as a basis for the ethical toolkit designed. The main phases are briefly explained next:



01 Deontological Stage

In the same way as the ethical framework for IoT, the framework proposed for AI begins with a deontological approach which is more an analysis of current legal frames regarding the topic. This stages helps to set a baseline for the goals and the vision of the project. An example of this stage would be to consider GDPR as part of the ethical strategy for AI, as well as the code of ethics that the Implementors or Researchers are subjected to.



03 Virtue Stage

This stage describes a step in the process intended to conclude in a code of ethics, moral code, or in simple "best practices" for the ethics of AI. This step concludes the whole process in a reflective manner setting a set of moral rules that can be included in future loops of the framework (as a "legal frame" for example).



02 Consequentialist Stage

A consequentialist stage is included in order to analyse and discuss the values, ethical goals, and vision of the project. This stage is intended to trigger discussion around the ethical consequences of the project again

Something that is worth to mention is that as different stakeholders have different roles in an AI application development process, the relevance and applicability of the ethical framework to each stakeholder may also be different. As shown in Figure 37, the ethical framework could be applied to each level of the responsible stakeholders with respective changes in the level of responsibility and processes followed.

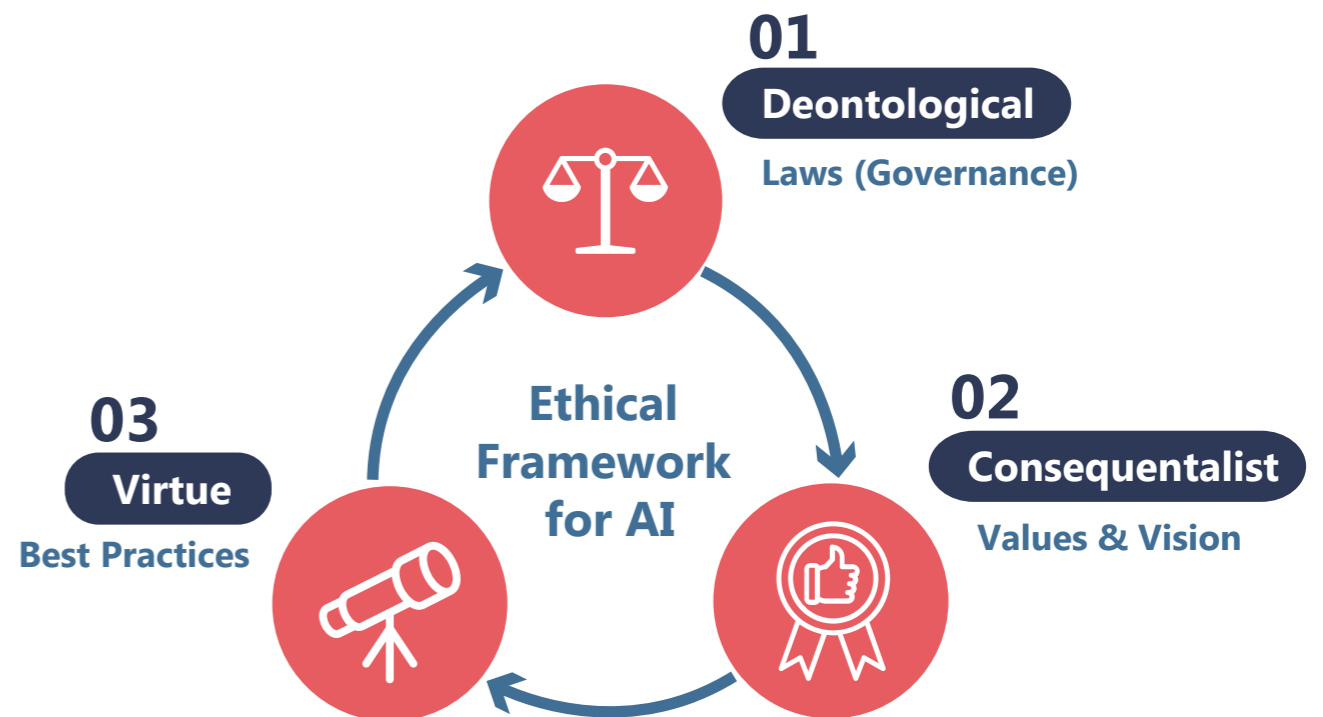


Figure 36: The Ethical Framework for AI developed in this section based on "The Ethical Composite framework" proposed by Seneges et al. (2017).

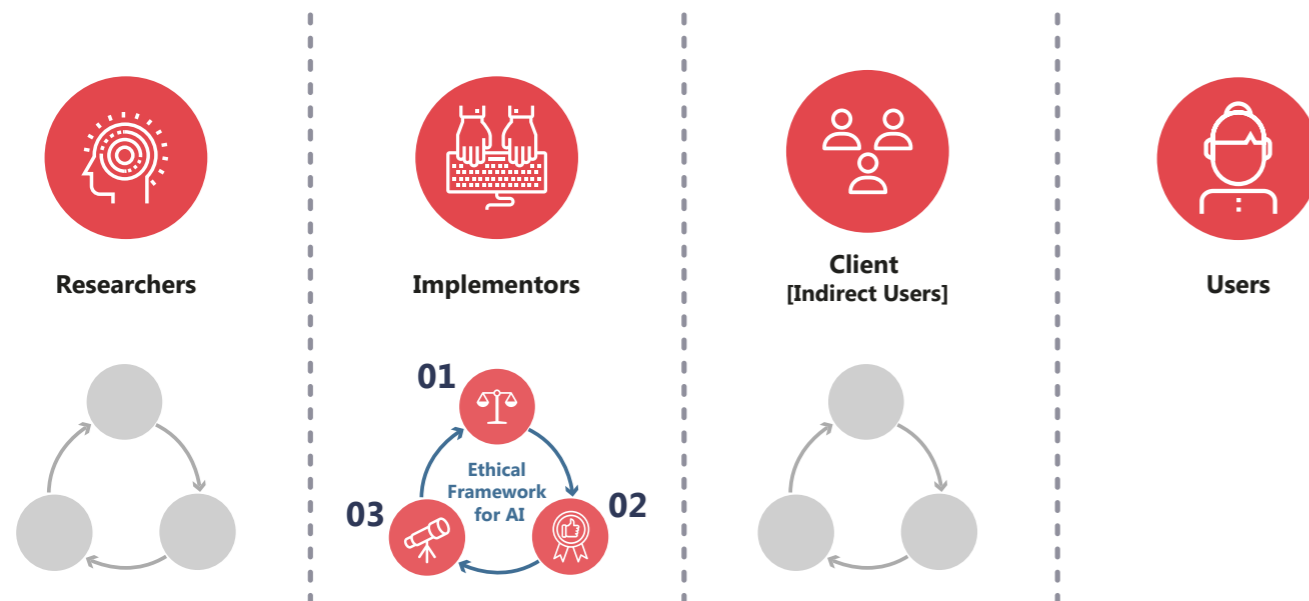


Figure 37: Visual overview of the application of the framework depending on the stakeholders' level.

5.3.3 An ethical canvas for AI applications

To explain the ethical framework to the company and other stakeholders, the ethical framework has been shaped into a canvas as shown in Figure 38. This canvas describes each element with a reflective question so it can be used for the analysis of the ethical considerations for AI, as some are enlisted in the canvas as well. The canvas works as a graphic representation of the three stages of any ethical framework (external interviews, Ethics of Technology researchers - TU Delft). First, a legal frame investigation is suggested. This frame could be external (regulations like GDPR) or internal (Code of Ethics or Good Practices). This is followed by the analysis of the vision and values that the developer or designer wants to reflect upon the product (brand vision and brand values as well). It ends up with a moral reflection that results from the ethical analysis and which is set into "good practices" or a "moral code". In the end, the "good practices" developed would become a "legal frame" and it could be iterated several times to encourage an ethical reflection and decision process.

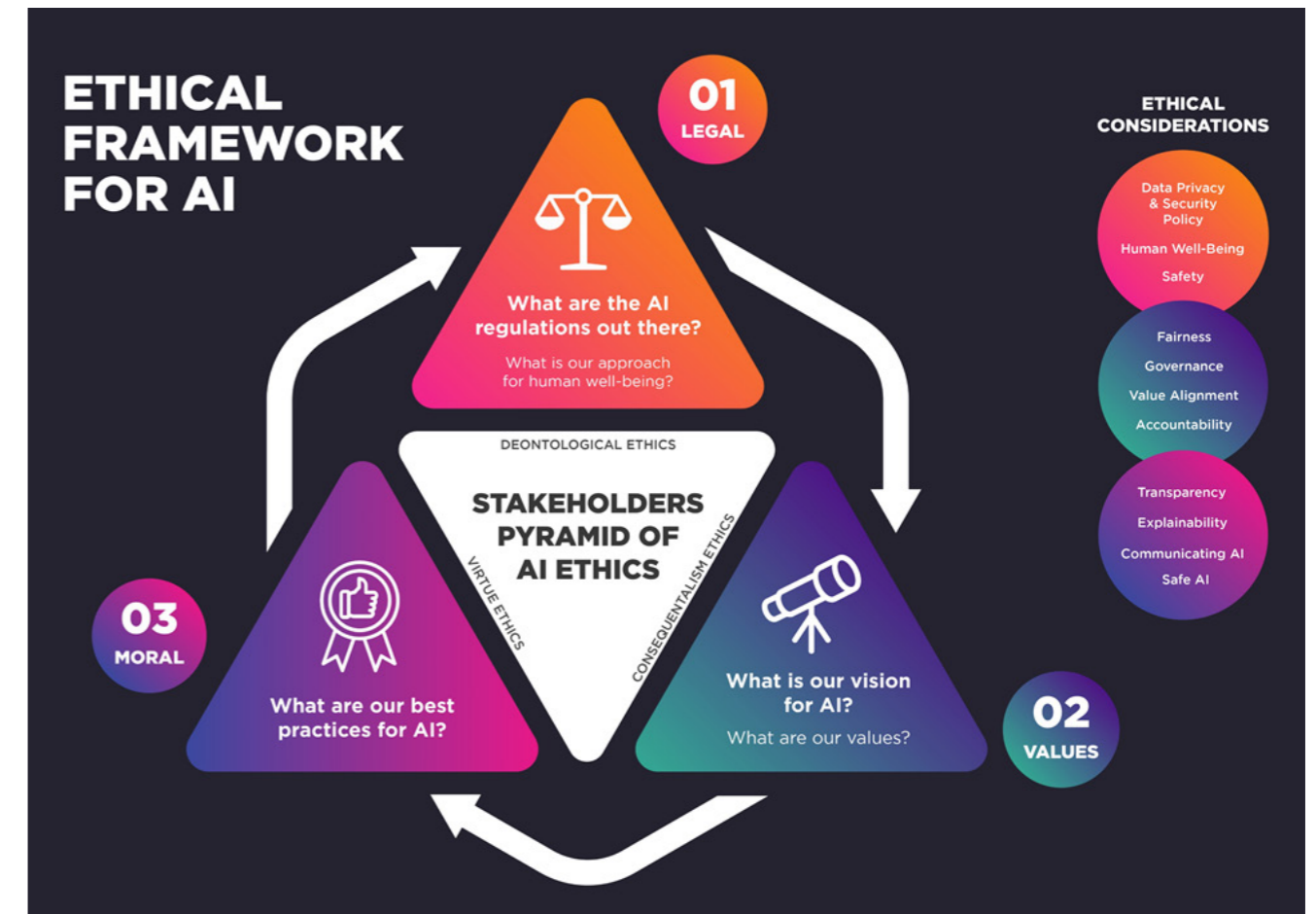


Figure 38: Canvas of the Ethical Framework for AI.

5.3.4 Discussion

In this section, a new ethical framework for AI was proposed based on a similar composite framework found in the literature. This framework was created using several ethical principles that were defined beforehand by doing ideation exercises. The framework features three ethical stages which are based on a deontological, consequentialist, and virtue approaches. The framework was translated to an actionable canvas in order to help with the understanding of the complex concept of ethics and its usage in AI systems development. This canvas features an "ethical cycle" that could be used to assist on the ethical evaluation of an AI-project. Moreover, the canvas also proposes the addition of the stakeholders involved in the development of AI applications which is proposed in the form of a metaphor.

5.4 THE PUZZLE OF AI ETHICS

This section describes the stakeholders layer that was added to the ethical framework canvas through an embodied metaphor of a puzzle. In this thesis, this puzzle metaphor is used to explain the importance of inclusion, diversity, and responsibility in the responsible development of an AI system.

5.4.1 Ethics is like a puzzle

In order to explain the framework and its additional layers in a relatable and playful fashion, a metaphor is proposed. This metaphor is one of an “ethical puzzle”, as it is intended to support the understanding of the complexity of ethics with a concrete concept (Saffer, 2005). This metaphor refers to the role of the stakeholders in the ethical process and the different levels of ethical responsibility that they have. On previous stages, it was found that the different stakeholders involved in the development of AI applications tend to work in “silos”, as they do not often interact with other groups of stakeholders (i.e. Researchers not consulting Users). As mentioned in Chapter 3, a development group formed by a highly diverse population would generate more ethical discussions, which would decrease the chance of ethical issues. This metaphor is intended to explain

the topic of ethics of AI with the inclusion of all the stakeholders involved, which is something that it is not entirely discussed in other ethical frameworks (Cutler, Pribic & Humphrey, 2018; Lubbock & Virdee, 2018), and which is important to consider for any ethical framework (Bonde & Firenze, 2013).

The selected metaphor tries to explain that the ethics of AI is like a puzzle, where you need to consider all the stakeholders involved in the process and build the puzzle with them to enhance ethics into your process. Moreover, this also shows the shared responsibility that all the involved stakeholders have over the development of an ethical AI application. The puzzle proposed has a pyramidal shape and it is composed of 5 different shapes that should be placed in a specific order to form the final pyramid. The pyramid then should be placed in a way to make the colors on each



Figure 39: The first prototype of the puzzle of AI ethics on top of the first version of the canvas of the Ethical Framework for AI.

face match with the ones in the canvas. As shown in Figure 39, the Researchers are featured in the base of the pyramid as they are the igniters of the technology. The pyramid climbs up towards the Implementors, Clients and the Users are represented on the top.

5.4.2 AR experience

In order to share the ethical framework with a large audience, as an internal and external marketing effort on the topic, a digital version of the canvas and the stakeholders pyramid was modeled. This model was rendered using the open source CAD program Blender. To make it available for more people, it was decided to upload the model to the AR Web platform 8th Wall. This interactive experience was shared with other stakeholders involved in the project like designers from FJORD, AI implementors and UX designers from MOBGEN | Accenture Interactive, and some professors from the TU Delft. The AR experience is shown in Figure 40 and can be accessed from a mobile device using the next link:

<https://8th.io/e3tq6>



Figure 40: AR marketing experience of the Ethical Framework for AI canvas.

5.4.3 Discussion

It is well known that metaphors support the understanding of complex and abstract topics by referring the intangible concepts to something more concrete. A “puzzle” metaphor would enhance understanding of the complex concept of ethics as well as generate attention to an often overlooked part of the development of AI, which are the stakeholders involved. Furthermore, the puzzle would also generate reflection towards the different levels of importance that each stakeholder has in the ethical development of AI systems, which is something that was validated on further testing sessions as described in the next section.

5.5 FRAMEWORK & CANVAS VALIDATION

This section describes the validation sessions made for the ethical canvas. Due to the involvement of a big diversity of stakeholders, the canvas was tested from different perspectives, including, experts in ethics, AI researchers, AI implementors from MOBGEN | Accenture Interactive and AI enthusiasts from Accenture Digital (Amsterdam).

5.5.1 Ethics of technology researchers

The canvas and the puzzle analogy were tested with an expert from the Ethics/Philosophy of Technology department of the TPM Faculty of the TU Delft. The tests consisted in letting the expert solve the “puzzle of ethics” individually and place the pyramid in the center of the canvas in order to encourage reflection about the role of all the stakeholders involved in the development process of AI (which are defined in previous chapters).

Gathered Insights

The framework is sustained in an Ethics of Technology model that conveys normative ethics theories in a “cycle of development”. During the test it was mentioned that this configuration brings both advantages and disadvantages since some of these theories could be used merely in a superficial way. The expert in ethics mentioned that the framework was clear and ethically correct as it worked as other ethical frameworks they. By starting with a deontological approach to look to any “legal or right frames” that already exist it would lead to the exploration of the definition of an ethical vision and values. Finally, the process will face a stage of reflection and the definition of moral practices (or moral code) that could eventually prevail as a “legal frame”, continuing the cycle.



Figure 41: Postdoc Researcher in Ethics of AI from the TPM Faculty of TU Delft testing the Ethical Framework canvas and puzzle (image by author)

5.5.2 AI researchers

Similarly to the validation stage with the Ethics of Technology researchers, another validation session was performed with a couple of researchers from the EEMCS Faculty of the TU Delft. The canvas and the puzzle analogy were tested this time and the tests consisted in letting the experts solve the puzzle together. This was made with the intention of explaining the advantages that collaborating and aligning for the same purpose have within the field of AI.

Gathered insights

The framework helped the researchers to understand that teamwork is useful when discussing ethical considerations of the development of AI. Furthermore, they found really useful to have a graphical guide for the discussion of the ethics involved in their research. They also praised the fact that the ethical considerations and principles for AI were listed inside the canvas. This, alongside the questions included in each step, allowed them

“It is interesting the way you fix everything together, the only thing that bugs me is the way you named each phase (as normative ethics approach). As an ethicists I could think about several objections for that...however, the process you propose is ethically correct. I think that this could open the discussion about ethics of AI within the organization you are working with.”

Postdoctoral Researcher - TPM Faculty TU Delft

to reflect upon what was important to add to their research in order to convey a more responsible outcome. Some points of improvement from their side were to include more questions in each step of the framework so a deeper level of ethics could be reached.

5.5.3 Company Tests

Several validation sessions were performed with key employees of MOBGEN | Accenture Interactive and also with a group of interns of Accenture Amsterdam. The sessions involved the discussion of the framework and specially the understanding of the “puzzle” metaphor of the canvas.

Gathered insights

In general, the feedback provided was positive, although, important insights were discovered during this stage. For example, some employees found the puzzle interesting but challenging enough, which allow them to expand their interest on the topic. On the other hand, other employees and interns found the puzzle too hard and frustrating. In the end, everybody agreed on that the puzzle adds not only a normally ignored layer of the development of AI but a playful manner to enhance interest on the topic of Ethics. Some improvement points were also discussed like the possibility of having an easier puzzle and instructions on how to use the framework included in the canvas.

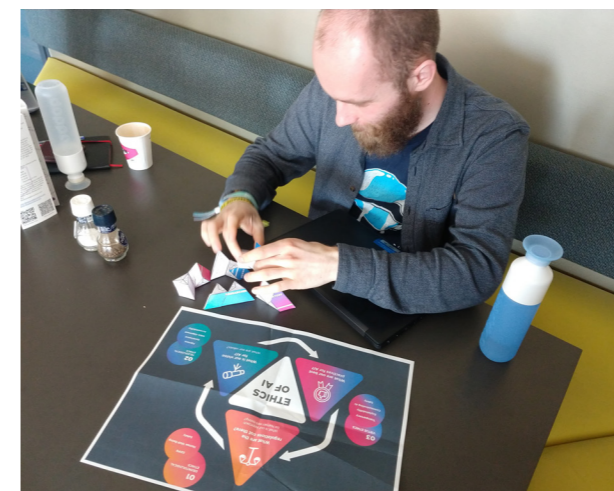


Figure 42: Internal test of the ‘ethical puzzle’ performed in Accenture Amsterdam HQ (image by author)



Figure 43: External validation session made with the help of researchers from EEMCS Faculty of TU Delft (image by author)

“I think that the values are the most important for us because...from this we can develop more ethical issues. To have all the ethical considerations here listed is helpful...”

Associate Professor - EEMCS Faculty TU Delft

“I liked the puzzle metaphor, it is nice to start the conversation around with colleagues, but to be honest it was really difficult to make it alone, I am not good at puzzles”

Postdoctoral Researcher - EEMCS Faculty TU Delft

“This is an interesting way of seeing ethics...but the puzzle is really hard”

Graduate Intern - Accenture Digital

5.5.4 Discussion

Certainly, metaphors are useful to explain complex concepts to people that is not familiar with a topic. In the case of this thesis, the use of a metaphor in the form of a “puzzle” was handy as most of the tests received positive feedback. The ethical framework was considered useful for most of the stakeholders involved in the validation sessions, although, constructive criticism was received as well. For example, that some sections were more useful for a determined group of stakeholders, or that the puzzle metaphor was interesting but at the same time it resulted in a hassle as some people are not good at making puzzles. On the other hand, the feedback collected was also used for the optimization of the framework and the canvas for future versions. It is important to mention that the canvas was validated by an expert in ethics of technology, which helps to bring academic credibility to the model.

5.6 DESIGN REQUIREMENTS FOR ETHICAL TOOLKIT

After the empirical studies and the development of the theoretical framework were performed, a set of design requirements for the ethical toolkit is defined to have robust guidelines to evaluate it and work upon. The requirements are based on the results obtained from the previously mentioned stages and are briefly described next:

- The design should be adapted (or adaptable) to the processes of MOBGEN | Accenture Interactive (or any other of Accenture’s dependencies). At the end of the day, the toolkit would be used by designers and developers from the organization.
- The solution conceived should incite critical reflection and trigger discussion about the topic visually and understandably. This way it is expected that all the stakeholders involved in the development of AI applications participate in the ethical and responsible implementation of the technology by understanding the abstract and complex topic of ethics.
- The ethical toolkit should contain a module or section that focuses on explaining the basics of ethics and its different approaches. This would solve the lack of ethical understanding and misalignment within the development teams.
- The final solution should be based on the theoretical framework developed, as it has been tested as a valid ethical evaluation process.
- The ethical toolkit should encourage the integration between MOBGEN | Accenture Interactive clients, designers, managers, and developers with other stakeholders impacted by the development of AI applications. This might be done directly or in a more indirect fashion (i.e. nudging to create awareness).
- Due to the stakeholders reach of the theoretical framework and the empirical studies, it would be advantageous if the ethical toolkit could be customizable for different contexts like academia as well as the AI development industry.

It is important to mention that the success of the project is not limited to the accomplishment of these requirements. Due to the lack of ethical assessment tools on the field and within MOBGEN | Accenture Interactive, it is expected that any contribution that improves the current situation around AI ethics would generate a considerably important impact.

5.7 CHAPTER CONCLUSIONS

In this chapter, the first research question was explored by transforming the insights gathered during the literature review and the empirical research stages to ethical principles for the development of AI systems. With this, an actionable framework around the ethics of AI was created. Furthermore, this ethical framework has been translated into a canvas that can be used by designers of MOBGEN | Accenture Interactive to trigger the discussion and evaluate the ethical considerations around an project that involves the implementation of AI. Additionally, a puzzle metaphor was introduced to generate understanding for people unfamiliar with the topic of ethics of AI. The coming chapter will explore the creation of ethical tools based on the framework defined in this chapter, intended for supporting the development teams of MOBGEN | Accenture Interactive and to enhance the critical innovation culture within the organization.

5.8 TAKEAWAYS FOR MOBGEN | ACCENTURE INTERACTIVE



The implementation of **ethical design methods within the process of the organization** (at all levels), could bring big benefits to the company.



MOBGEN | Accenture Interactive might use the provided framework to **explain to their clients the importance of ethics in the development of AI applications.**

06

chapter 6

Design Principles for the Ethics of AI

This chapter elaborates on the design process of the ethical toolkit for the development of AI applications. The design is based on the ethical framework explained and the design requirements defined in the previous chapter. First, an internal research was performed using semi-structured interviews, generative exercises and a small survey in order to understand the processes followed and ethical views inside the organization. Then, a couple of “provotypes” were used to create an awareness internal campaign on the topic. Finally, using all the insights gathered, an iterative design process was followed to create the toolkit.

6.1 INTERNAL INTERVIEWS

This section describes the internal research stage performed within MOBGEN | Accenture Interactive, made by a series of interviews with important stakeholders of the company. It elaborates on the insights collected, specially on how the company works and how it might incorporate a responsible AI methodology in its processes.

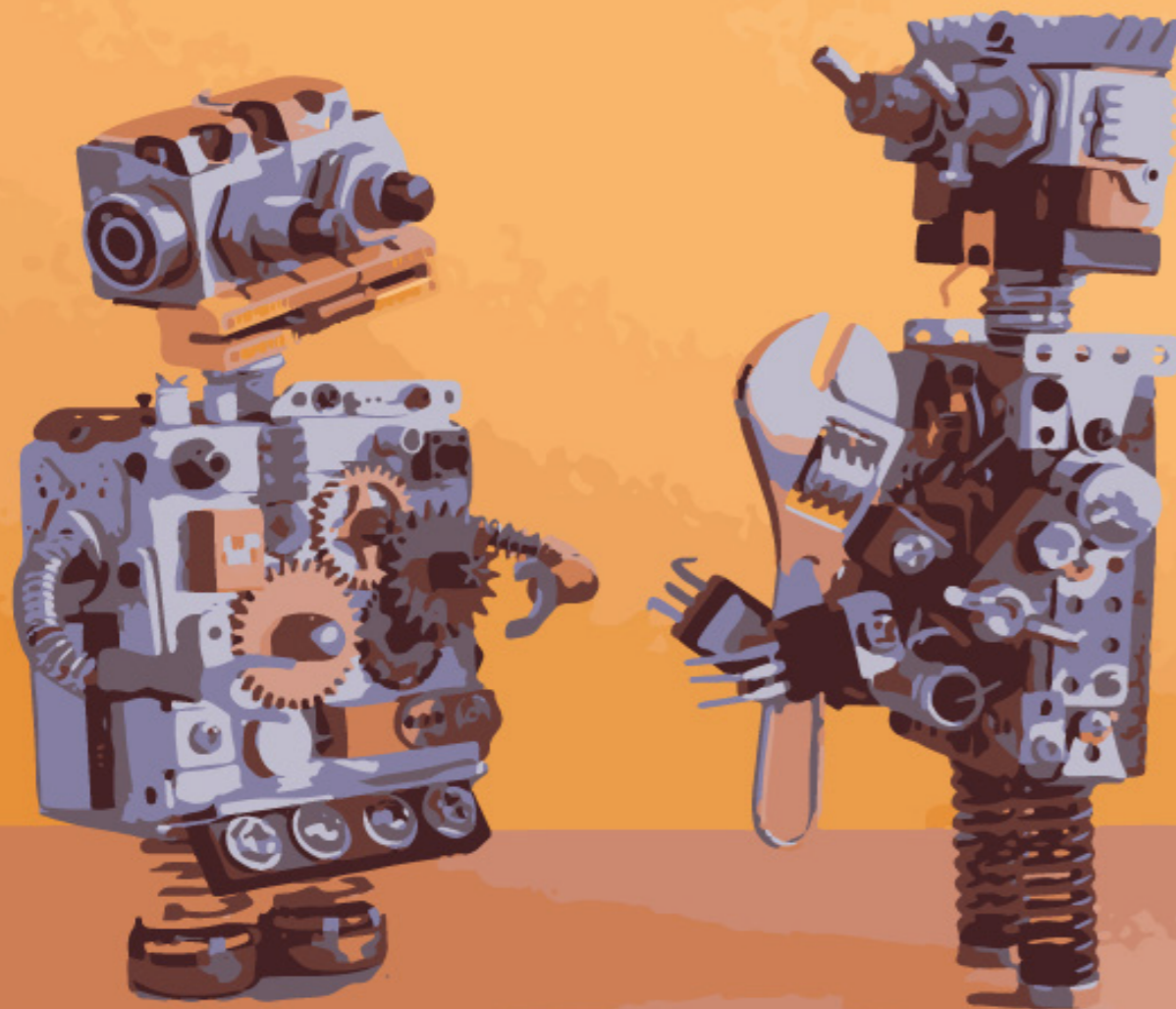
6.1.1 Knowing MOBGEN | Accenture Interactive better

Six semi-structured interviews were conducted to understand the current methods and processes followed by the company. Most of the interviewees hold a managerial position related to the digital development and design departments of the company. The main goals of this exercise covered:

- Defining the added value proposition that MOBGEN | Accenture Interactive offers to their customers.
- Discovering (if any) the type of ethical assessment that is performed during projects.
- Understanding the processes followed by the organization at different levels (high, medium,

and low levels).

- Finding the way of integrating an ethical approach or method to the company's development processes.
- Looking for areas of opportunity in other Accenture dependancies like FJORD.



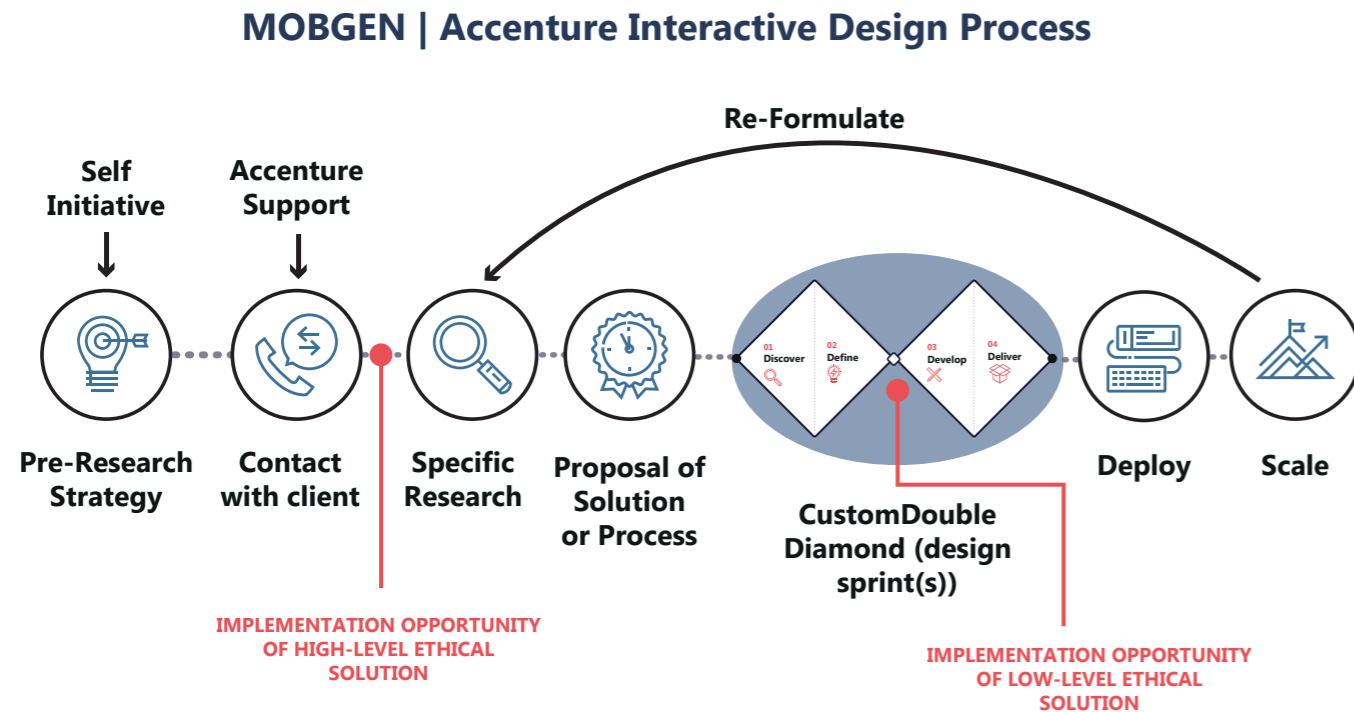


Figure 44: Visual representation of the design process followed within MOBGEN | Accenture Interactive. The illustration also shows where in the process, an ethical assessment could be implemented to generate value to a project, specifically about AI (illustration by author, 2019)

6.1.2 What was learned?

The insights collected from the interviews include the different processes that are followed within the company at different levels. The general process is divided in several stages that go from a research stage to a scaling strategy at the end. MOBGEN | Accenture Interactive includes a custom double diamond approach within the process which is followed for the development of most of the digital projects within the company. It is important to mention that this sub-process follows an agile approach as it is sometimes performed in a “design sprint” setting. Some other insights include the lack of an ethical assessment to the design brief constructed with the help of the client, as the only discussion about ethics that takes place in a project is when its intention generates unconformity among the development team, as was also identified in Chapter 4. This information helped to define where in the process an ethical assessment

stage could be performed and how it might impact the project at a higher or lower level. Furthermore, the interviews also included the input from other Accenture Digital entities like FJORD.

One of the most interesting insight from those interviews was that ethics could be perceived more as an opportunity instead as only a negative risk if the company does not comply with it. To sum up, because of the insights discussed previously, and due to the poor involvement of the topic within the development cycle, it was concluded and decided that the solution (or solutions) designed would try to cover all levels of decision making within the company. Moreover, the ethical toolkit should try also to trigger an “opportunity” notion of ethics, something that has already been mentioned in previous academic research (Boddington, 2017).

“There is a lot of thinking on a political level on what you do, retraining...that’s not our place (...) but for a process if we would introduce an API or a RPA for a company that’s got a call center of 3000 people and reduce it to 1500, then we need to be aware of what means from a human cost.”

CCO - MOBGEN | Accenture Interactive

“I think that you could try to see ethics less like a risk and more like an opportunity. It is always framed like something negative, when it can give you options to innovate.”

Systems & Service Design Lead - FJORD The Dock

“Before looking if an AI application is ethical, we look to see if it is serving a purpose”

Product Lead - MOBGEN | Accenture Interactive

6.1.3 Discussion

This section describes the insights gathered about some of the internal processes followed by MOBGEN | Accenture Interactive. Six semi-structured interviews were performed where the added value proposition that MOBGEN | Accenture Interactive offers to their customers was discussed and defined. The organization follows a custom double diamond process with agile sub-processes that form their widely known design-sprint. It was also confirmed that there’s no ethical assessment of any kind during projects, and if there is one (i.e. GDPR), it is managed by a high level and a legal department as “compliance” only. It was also concluded that the solution designed would try to cover all levels of decision making within the company. Moreover, in order to look out for more tacit or latent needs for employees, a generative exercise was proposed to make with a couple of key employees as it is discussed in the next section.

6.2 GENERATIVE EXERCISES & SURVEY

This section elaborates on the generative exercises and online survey used for the collection of insights regarding the current processes followed and the ethical knowledge within the company.

6.2.1 Generative exercises

Using some of the insights of the previous stage, a couple of generative sessions were organized with key decision makers of MOBGEN | Accenture Interactive. An account manager and a service design lead provided relevant information by doing a stakeholder mapping and a graphical description of the process or processes followed on their roles.

Gathered Insights

The generative exercises complemented the information gathered from the internal interviews. Relevant stakeholders, and its hierarchy within the company, were defined thanks to the stakeholder mapping exercise. Moreover, the description of the process followed by the participants resulted in interesting insights regarding the ethical views and practices of the company. For instance, GDPR considerations are really important for both, the company and its clients. On a higher level, this considerations are managed and specified by a legal department before starting any project. Once this happens, designers and developers must make sure that these regulations are covered throughout the project. This shows that the approach to ethics within the company still follows a “compliance” path for ethical issues, instead of a proactive role towards it. This strengthens the need for tools that encourage ethical discussion of other relevant ethical principles that involve AI, besides data privacy, among the development teams. In order to complement the information obtained until now, a short online survey was performed as described next.

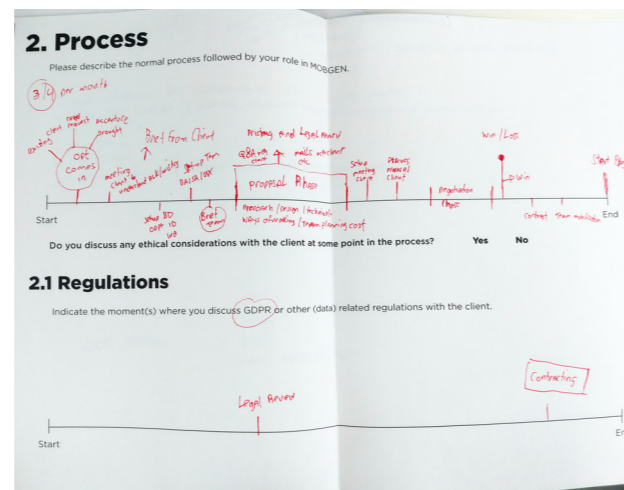


Figure 45: Sample of the generative exercises filled by MOBGEN | Accenture Interactive employees.

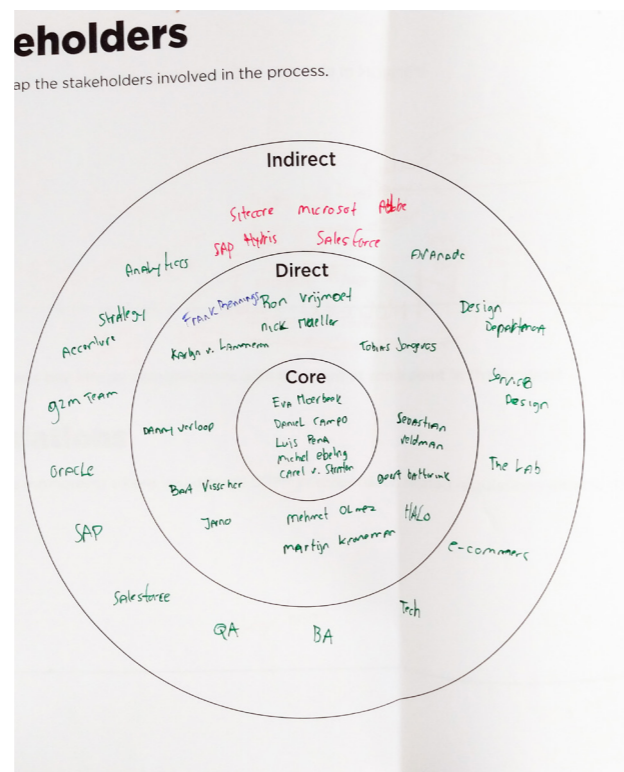


Figure 46: A stakeholders map (for a given project) created by an account manager from MOBGEN | Accenture Interactive. Blue colored text represents a high hierarchy in the company, green a low hierarchy, and the written in red are external partners.

“I don’t think that our clients have enough maturity in the topic to understand the ethics of AI”
Account Manager - MOBGEN | Accenture Interactive

6.2.2 Online Survey within MOBGEN | Accenture Interactive

An online survey was distributed within the company in order to understand the ethical culture of its employees. The survey featured 10 questions that inquired about the perceived ethics of the company as well as the ethical views of its employees. Some questions also referred to the ethical strategies followed by employees and their participation in AI related projects. A detailed view of the questions and results obtained are shown in Appendix E.

Gathered Insights

Unfortunately, the amount of respondents was not enough to draw a clear conclusion, however, interesting insights arose. These are discussed next:

- Most of the surveyed employees understand the term Responsible AI as a robust term (bigger context), although, some respondents considered the term is only related to the ethics of AI in particular.
- It was stated that employees do not follow regularly a type of ethical assessment (i.e. reflection session, ethical discussion with client, etc) on a project, which is something that already have been observed and discussed during the internal interviews stage with managers and employees of the organization.
- One of the most interesting insights found is the fact that almost half of the surveyed employees do not know about the ethics policy of the organization (Code of Business Ethics), something that is rare as there is a mandatory Ethics Training that all employees must do during the first months of employment.
- Complementing this, the majority of the respondents confirmed that they have considered other ethical implications besides GDPR during a project.

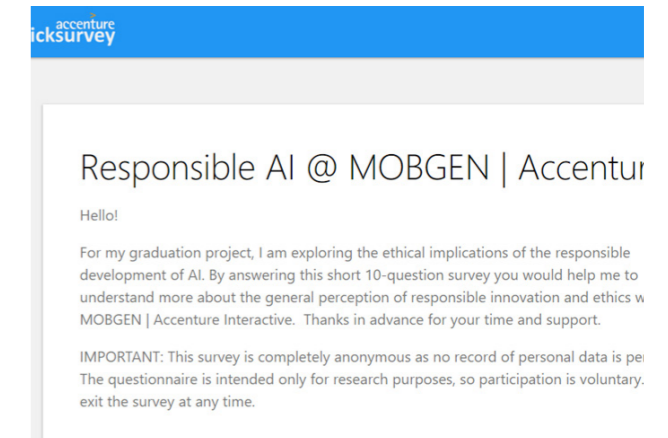
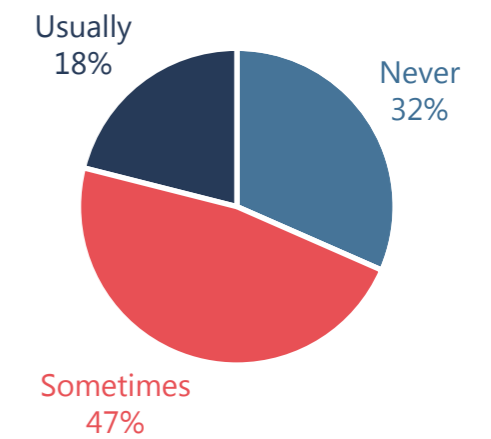


Figure 47: Introduction Message of the quick ethical survey.

How often do you follow any type of ethical assessment? (i.e. reflection session with team, discussion, etc)



Does your organization have a written ethics policy?

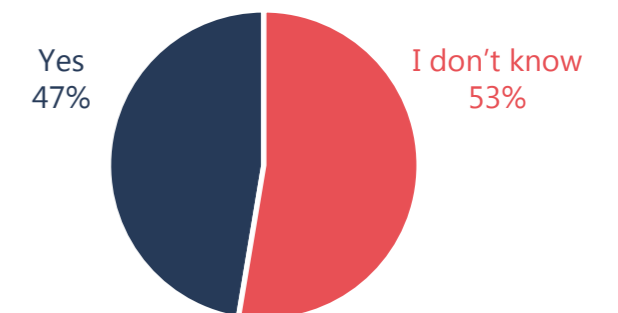


Figure 48: Sample of some of the results of the quick survey

This exercise demonstrated that there is a reasonable interest from employees towards the topic of Ethical AI, however, ethics in general is still taken as trivial. It was observed that ethical assessment is not performed or performed rarely. Furthermore, it was also shown that employees follow their own ethical approach while doing a project, considering that GDPR considerations are looked upon continuously for compliance. Certainly, they might follow their own ethics without knowing or taking into consideration other ethical views. This is something that has already been observed in previous research (Chapter 3) and that has shown that it could provoke more ethical issues in the future (lack of diverse thinking). The creation of ethical frameworks and tools, would help employees to be more aware of the topic and have the opportunity to assess and reflect upon their ethical decision during the creation of any AI application.

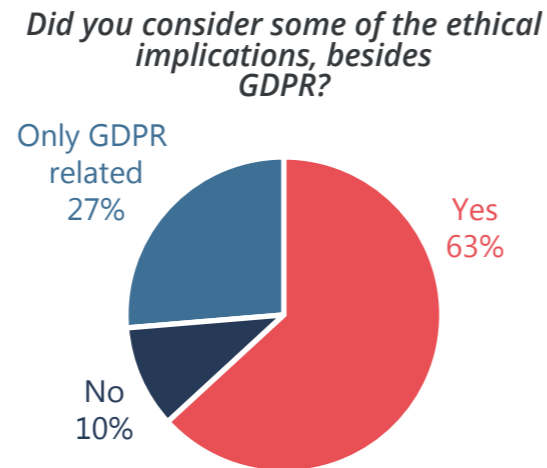


Figure 49: Sample of some of the results of the quick survey

6.2.3 Discussion

A complementary stage of research on the company processes and internal ethical views was elaborated upon in this section. The generative exercises complemented the information gathered from the internal interviews as the short ethical survey showed that there is an interest on the topic by most of the employees of MOBGEN | Accenture Interactive, although, ethics is still taken as a trivial topic. One of the most relevant findings of this stage is that the approach to ethics within the company still follows a “compliance” path, instead of a proactive role towards it. Moreover, using the short survey it was discovered that employees follow their own ethical approach while doing a project, considering that GDPR considerations are looked upon continuously for compliance. Because of this attitude towards ethics, an awareness campaign was planned via an action research cycle as shared in the next section.

6.3 ACTION RESEARCH CYCLE 2 - IMPLEMENT

Similarly to previous phases, an action research cycle was executed during this stage as well as described in this section. This cycle featured two activities that were intended to generate awareness and to trigger the attention of MOBGEN | Accenture Interactive employees to the topic of Ethics of AI. First, an ethical awareness campaign was performed using a prototype, with the intention of giving the project some notoriety within the company as well as to generate concerns and awareness on the topic of unfair AI. This activity would also provide an overview of the perceived importance that the topic has within the organization, as well as show the individual ethical views of employees. Furthermore, the second activity featured a “guerrilla marketing” campaign that exploited a “failed” internal marketing campaign. This activity created interest in the topic as the effort was considered “bold” by some employees. The insights from this stage assisted the iterative ideation of the ethical tools that was performed afterwards and is discussed in the next section.

6.3.1 Implementing awareness via Prototypes: “Bias in Coffee Machine”

As part of the internal awareness campaign around the topic of Ethics of AI, a “prototype” that played with the idea of AI bias was developed and installed in the coffee machine of the company. This prototype was made as a tangible strategy in the form of an interactive “wizard of oz” application. The whole scenario embraced the idea of bias within AI decision-making and how this unfair moment was considered unethical.

The prototype was developed as an Android application that featured a “badge scanner” to encourage employees to interact with it. The application aimed to trigger awareness by provoking confusion, surprise among the employees. The application started with a provocative notice regarding the current free coffee being charged from a certain point due to company policies. The application then asked employees to scan their company badges in order to notify them if they



Figure 50: Diverse screens of the ‘prototype’ created for this Action Research Cycle.

were allowed with an exemption. The application was coded to provide a random binary answer. The experience ends up with the presentation of questions about the fairness of the decision and the importance of ethics while designing AI applications.

The location and topic of the provotype were selected using design for intent techniques (Lockton, 2010). Design for intent aims to give designers a more detailed approach to behavior, understanding, and the complexities of everyday life. It also explores the interactions between design and people's behaviour, across products, services and environments.

Mostly positive feedback was collected from the provotype intervention, as well as comments about the awareness of the topic triggered by it. The experience also opened a thread of discussion regarding the topic of ethics along several internal communication channels. Figure 49 shows that the objective of the provotype was achieved by making employees engage in ethical discussions. This discussions also display the general view of the company towards the topic and the need of tools that help employees understand the different ethical approaches to a certain situation. This is an interesting insight that was taken into consideration during the iterative ideation of the ethical toolkit.



Figure 52: Setup of the provotype installed on top of the coffee machine.



Figure 51: The provotype sparked various reactions from employees as they also provided valuable feedback.

"I think that the values are the most important for us because...from this we can develop more ethical issues. To have all the ethical considerations here listed is helpful..."

Associate Professor - EEMCS Faculty TU Delft

"I liked the puzzle metaphor, it is nice to start the conversation around with colleagues, but to be honest it was really difficult to make it alone, I am not good at puzzles"

Postdoctoral Researcher - EEMCS Faculty TU Delft

"I liked the puzzle metaphor, it is nice to start the conversation around with colleagues, but to be honest it was really difficult to make it alone, I am not good at puzzles"

Postdoctoral Researcher - EEMCS Faculty TU Delft

6.3.2 Awareness Provocative Posters

To complement the awareness campaign, a series of provocative posters were placed inside one of the buildings of the company. By taking advantage of a new internal marketing campaign, the provocative poster aimed to remind the employees of the importance of the Ethics of AI. Its message related

to the ethical importance of context when designing an AI application. The posters were removed one month later after they were placed. This shows that the poster helped to make the topic relevant within the organization.



Figure 53: A sample of the guerrilla marketing strategy used to promote the project within the company (picture by the author).

6.3.3 Discussion

The exercises described in this section generated awareness and triggered the attention of MOBGEN | Accenture Interactive employees to the topic of Ethics of AI, as shown by the feedback gathered. The ethical awareness campaign performed via a provotype and the "guerilla marketing" poster got some attention from employees as well as from high level management. These exercise brought interesting insights and results like a discussion around the topic within the internal communication channels of the company. Nevertheless, although these actions triggered discussion within the organization, it was also noticed that the topic was still complex and new for MOBGEN | Accenture Interactive. Because of this and considering that the solutions should aim to assist the development teams, a more clear structure was required. This was made by translating the design insights into a strategic vision, which is shown and analyzed in the next section, and that also helped during the iterative ideation of the ethical tools.

6.4 STRATEGIC VISION OF THE PROJECT

In order to help MOBGEN | Accenture Interactive to add value to their development process with an ethical value proposition, a strategic vision was defined. The strategic vision defined alongside the ethical framework previously explained, served as a basis for the ideation of the ethical toolkit.

6.4.1 A strategic vision for the Ethics of AI

The vision formulated in this chapter defines the reasoning behind the design of the ethical toolkit. It takes into account the insights gathered through the literature review stage as well as the empirical interviews and the first sections of this chapter. It is proposed mainly with the objective of framing the design principles that the ethical toolkit should follow to create value for the company.

01 Ethics as an opportunity, not only as compliance

This thesis aims to change the understanding of ethics within the organization from a “compliance only” perception to fuel for innovation in the development of AI applications. This new notion aligns with the suggestions made by designers during the research stages, where a “less negative, more opportunity based” approach towards ethical outcomes was discussed (i.e. internal interviews). More importantly, this new notion of ethics would generate an added value for MOBGEN | Accenture Interactive by helping their clients to expand their use of AI and at the same time allow them to be aware of their accountability and the unintended ethical consequences of this implementation.

Furthermore, ethics implementation could promote new opportunities for innovation and avoid consumer trust crisis that affects both big and small corporations alike nowadays.

02 Ethical Building Blocks

Proposed by Simons (April, 2019), the building blocks required for organizational capacity to support ethical outcomes are included in the strategic vision. These blocks follow the premise of ethical outcomes by having an ethical organization, which at the same time provides ethical processes and tools to create ethical people or employees. A visual representation of the blocks are shown in Figure 54. This logical process aligns with the objectives of this thesis as the final solution would concern an ethical tool that aims to create ethical employees.

03 Context always matters

It can be stated that ethics depends on a determined context where a set of moral values is defined. Because of this, and due to the fact that context is vital for the strategic design practice, context of the implementation of an AI application should be explored and revisited as often as possible in order to have a more ethical practice.



Figure 54: Ethical Building Blocks adapted from Simons (April, 2019)

04 Increase ethical reflection and alignment

As mentioned continuously in previous sections, designers and developers often use their own ethical judgement to solve delicate situations during the development of AI applications. Moreover, the organization hardly enhances the coordination of reflective sessions within a project. Hence, this thesis states that by increasing and encouraging reflection within the project, a more ethical AI implementation would be performed in the long run.

6.4.2 Target group

From the literature review, empirical research and the internal research performed we can conclude that there is a need for more ethical people in order to create an ethical AI applications development. Furthermore, ethics in the corporate world is closely related to brand trustworthiness, which is a vital part of any organization in today's global interconnected economy. In simple words, an ethical brand is a trustworthy and a reliable brand (Singh et al., 2012) . Because of this premise, the target group of the strategic vision is the employees of MOBGEN | Accenture Interactive that participate during the development of AI applications. This teams generally consists of: Account managers, design and development managers, development leads,

design leads, creative technologists and service designers. It can be observed that a large diversity of different stakeholders participate during an AI implementation project, including different levels of hierarchy. Because of this, a solution that supports all levels is preferred.

6.4.3 Discussion

A strategic vision and its target group of are analyzed in this section in order to help MOBGEN | Accenture Interactive to add value to their development process. This vision defines the reasoning behind the design of the ethical toolkit as it helps framing the design principles that the ethical toolkit should follow to create value for the company. The vision is based on four main pinnacles: ethics as an opportunity, strategic ethical building blocks, context always matters, and increase ethical reflection and alignment. The vision is intended to assist the employees of MOBGEN | Accenture Interactive that participate during the development of AI applications, which generally consists of: Account managers, design and development managers, development leads, design leads, creative technologists and service designers. This vision is one of the main pinnacles of the developed toolkit which was iteratively ideated as presented in the next section.

6.5 ITERATIVE IDEATION

This section elaborates on the ideation process followed for the development of the ethical toolkit. This process was performed iteratively and it relied on the assistance of several stakeholders during validation and test sessions. These sessions are described more in detail in Chapter 8.

6.5.1 Ethical Strategic Blueprint

Need:

Within MOBGEN | Accenture Interactive there is a lack of general ethical assessment while developing an AI implementation project. This may lead to an ethically misaligned AI application, which would affect the value proposition of the company and its brand image. For instance, it has been documented that due to plenty recent ethical scandals, Facebook's brand is now catalogued as untrustworthy (Francis, 2017; Zhang & Dafoe, 2019). Hence, a design strategy in the form of an applicable workshop is designed to add value to the MOBGEN | Accenture Interactive and involve ethics in their current AI apps development proposal. This blueprint was developed based on the Ethical Framework for AI proposed in Chapter 5 as seen in Figure 55. Most of the developed tools were ideated personally, however, both The Evil of [AI] game as well as the Responsible AI Deck were developed with the support of internal and external stakeholders.

Research Insights:

This setting was chosen as a first effort to cover the ethical considerations for the development of AI applications to create an added value proposition for the company. The lack of current solutions that involve ethics in the development process is evident. Moreover, the disperse and unaligned understanding of ethics within the company strengthens the notion for a general strategy. Therefore, this workshop is intended to not only add value through the eyes of the company's clients, but also by providing support to their employees in an effort to encourage a wider ethical judgement (ethical building blocks).

An Ethical Strategic Blueprint:

The chosen configuration to make the strategy tangible and useful for the company is a full-day workshop setting. This workshop would feature two important ethical stages. First, it is important to create an uniform understanding of ethics and its main approaches. Because of this, an "Ethics & [AI]" mini-workshop is proposed. This workshop concludes with the implementation of a game called "The Evil in [AI]" that aims to trigger ethical alignment within the development team and, at the same time, enhance discussion around the topic of ethical consequences of AI. This first stage is recommended to be performed alongside the client, making everybody aware of their ethical views and enhance an alignment on the ethical understanding of the project.

In addition to this, a second stage where the ethical aspects regarding the development of the project are explored and addressed. This stage begins with the creation of a project overview checklist by the development team. It is followed by the ideation of ethical risks and opportunities by a set of trigger cards. It ends with the identification and mapping of the risks for its evaluation on an Ethical Axis. Finally, the strategic blueprint would end with away of translating the main insights of this analysis to ethical specifications that should be followed during the rest of the project.

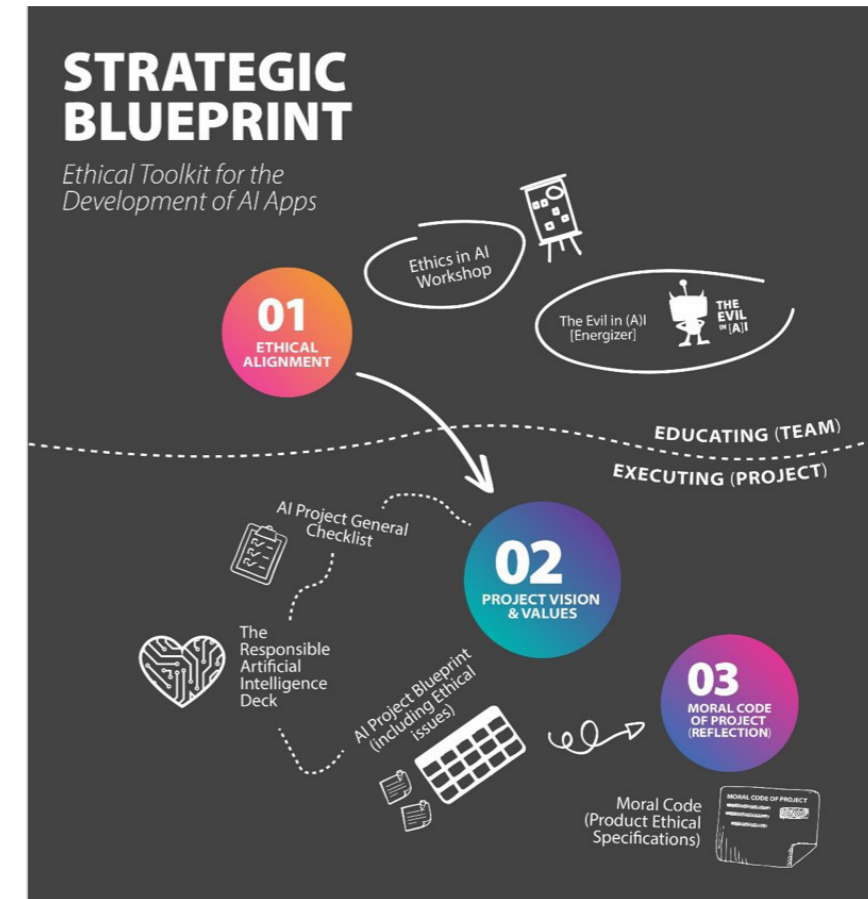


Figure 55: Overview of the Strategic Blueprint included in the Ethical Toolkit

6.5.2 Ethical Mini-Workshop & Game

Need:

Several needs were identified that resulted from the previous research stages were addressed during the ideation of the first stage modules:

- **Create understanding and alignment on the concept of ethics and its different approaches.**
- **Enhance familiarity with ethical decision making, which would define the ethical approach of the project.**
- **Explore the importance of the societal consequences of unethical AI applications.**

"Ethics & [AI]" Mini-Workshop:

As mentioned before in this chapter, one of the value propositions of MOBGEN | Accenture Interactive is the organization of workshops for and with the participation of customers. Hence, a workshop

setting to explain ethical concepts was selected due to the familiarity that the organization has with the technique. Furthermore, it has been demonstrated in previous research that a workshop session that enhance participant activity and encourages participation has positive effect in the professional practice (Davis et al., 1999; Halskov & Dalsgård, 2006; Sanders & Stappers, 2013). Two iterations were made for the creation of the "Ethics & AI" Mini-Workshop. The design features a presentation aimed to explain a general overview of ethics, its relevance in the context of AI and its different approaches. The workshop includes also a section that explains the importance of the ethical decision making and why ethics is a complex non-binary term. This is an important objective for MOBGEN | Accenture Interactive as they have to show leadership towards the topic of responsible innovation to differentiate from other ethical strategies in their current market

by including also an ethical educational section in their value proposition. Moreover, this feature would help their clients improve their market value via the acquisition of knowledge on the different concepts of the ethics of AI.

A validation of the first version of the workshop was performed during an “experience chapter” meeting organized within MOBGEN | Accenture Interactive and it is described more in depth in Chapter 8.

The Evil in [AI] game:

As previous research demonstrates, gamification has proven to be very successful in enhancing engagement and behavioral changing (Burke, 2013; Kumar, 2013; Hamari, J., Koivisto et al., 2014). Because of this, and to promote discussion and support a critical innovation process towards AI applications, a game was ideated as a good fit for the end of the first stage of the ethical strategy. The Evil in [AI] exercise is a game designed to explore the importance of the societal consequences of unethical AI applications. It aims to trigger discussion and reflection about the most evil AI system that the participants could create, in an original and playful manner. A first iteration of the game was made during a validation session with a couple of AI researchers from the TU Delft. The insights gathered from this session could



Figure 56: Sample pictures of the first version of the Evil in [AI] game.

be consulted in Chapter 8. The mechanics of the game have been changing in order to enhance the approach to the need addressed. The first version of the game is shown in Figure 56 and Figure 57. A more updated version and the final mechanics of the game can be consulted in Chapter 7.



Figure 57: Overview of the first version of the Evil in [AI] game cards.

6.5.3 A Responsible AI strategy for the project

Needs addressed:

The second stage of the ethical strategy proposed covers the needs identified around ethical assessment and tools to incorporate it into the design sprint that the company performs with its clients. First, a General Checklist is proposed with the objective of making the team understand the ethical impact that the project could have, as well as to create awareness of the ethical principles that should be considered throughout the project. This overview of the project should remain visible and present during the project to support ethical decision-making. Then, certain ethical risks and possible opportunities are identified with the support of “trigger-cards”. Finally, these risks (opportunities) are analyzed and mapped among an ethical evaluation axis that measures the ethical impact of the risk along the probability of happening. This stage addresses the need of incorporate ethics into the design process of the company.

AI Project General Checklist:

The General Checklist of the AI project is a engaging way to trigger a discussion about the new AI application. Designed as a poster-size canvas, it is intended to make the development participants aware of the implications, impact, and ethical

interaction that the projected app would have. Furthermore, it provides a graphical overview of the characteristics of the project.

Responsible Artificial Intelligence Deck:

This strategic solution follows a “trigger card” configuration. This type of cards are created to support design and development teams with related and (sometimes) unforeseen topics to spark their imagination and problem solving skills. This deck was ideated taking into account the insights obtained from the literature review as well as from the empirical research studies. This configuration was chosen, over a classic “specific scenario proposal” (i.e. TriggerCards by Alejandro Masferrer), due to its generality of the topic and easy adaptation to different types of users. For example, the cards cover the topic of Ethics of AI from a general perspective providing 4 aspects to consider, including the stakeholders impacted, the context of the intended AI application, as well as several ethical principles defined previously. Some support cards containing Human Values were added in case the quest for ideas gets to a dead end. This configuration allows any user to have a tangible and visual perspective of the ethics involved in a project. Or even better, other usage methods ideated contemplated the idea of using the cards just as a conversation starter in client meetings or as support material for the creation of a new AI application from scratch (opportunities). A first iteration of the cards is shown in Figure 58.



Figure 58: Prototyping process of the Responsible Artificial Intelligence Deck



Ethical Evaluation Axis:

This evaluation matrix axis lets you quickly prioritize a large list of potential ethical consequences for the project, in order to prioritize the course of action during the project. This exercise helps to determine the most impactful and most likely ethical risk that the team should act on. Visualizing the prioritization of the identified ethical risks facilitates an ethical conversation within the development team. Moreover, this activity supports also the team with the decision making of the ethical considerations of the project.

6.5.4 The Moral Code for an AI Application

In order to transform the ethical risks and areas of ethical challenges to product specifications, an ethical specifications sheet is proposed. Following the third and last stage of the Ethical Framework, it asks for the ethical considerations of the project in hand. These go from a scale of Always or never OK ethical situations to a middle ground of acceptable if it meets certain conditions.

MORAL CODE OF THE PROJECT:

It's always OK to _____	It's acceptable to _____ if _____	It's unacceptable to _____ if _____	It's never OK to _____
😊	😊	😞	😞
NOTES:			

Figure 59: Sample of the first version of the Moral Code.

6.5.5 Discussion

This section describes the aspects considered during the iterative ideation of the modules for the ethical toolkit. The ethical toolkit was ideated based on a strategic vision and an ethical framework developed during the project. The toolkit follows a strategic blueprint that features two different stages for the development of an ethical perspective towards AI. The first stage is formed by two modules, an "Ethics & [AI]" mini-workshop and an energizing game called the "Evil in [AI]". The main objective of this stage is to create a uniform understanding of ethics and its main approaches and to trigger and support dialogue about the ethical consequences of AI. The strategic blueprint continues with a second stage that aims to provide ethical tools for the design teams that are involved in an AI related project. This stage follows an ethical evaluation scheme that concludes with the generation of a moral code for the AI project. The aim of the second stage is to identify the ethical risks and areas of ethical challenge and to enhance an ethical discussion among the participants involved in the project.

6.6 CHAPTER CONCLUSIONS

This chapter shows the methods used to know more about the organization and how to adapt the ethical tools to its development process. From the interviews performed, the current process followed within MOBGEN | Accenture Interactive was described and it was concluded that the solutions ideated would try to cover all levels of decision making within the company. Moreover, the ethical toolkit should try also to trigger an "opportunity" notion of ethics, instead of the "compliance" role it is always given. Additionally, a couple of generative exercises were executed as well as a short ethical survey that resulted in insights like the way MOBGEN | Accenture Interactive deals with ethical concerns like GDPR and the general ethical knowledge and views of some employees. A second action research cycle was performed with the intention of create awareness and interest on the topic of ethics of AI inside the company. The observable results were successful as the feedback received was positive and a thread of discussion was started in the internal communication channels of the organization. These activities were vital for the ideation of the ethical toolkit discussed in the final section of the chapter. An ethical strategy was devised and translated to an strategic blueprint, which features two stages. The first stage is formed by 2 modules and its main objective is to educate the development teams and the client by creating a uniform understanding of ethics and supporting dialogue about the ethical consequences of AI. The second stage is formed by 6 modules that aim to provide ethical evaluative tools for the design teams involved in AI projects. The final version of the designed ethical toolkit is described in Chapter 7 and all the validation sessions performed to test the toolkit are shown in Chapter 8.

6.7 TAKEAWAYS FOR MOBGEN | ACCENTURE INTERACTIVE

MOBGEN | Accenture Interactive should **consider to develop internal experiences around Accenture business ethics in general and the ethics of AI** in order to make employees more aware of the ethical regulations of the company.



The organization **might consider to adopt provotypes into their value proposition** as part of the experiences they design.

chapter 7

An Ethical Toolkit for AI

This chapter describes the final outcome of this project. The final design is an ethical toolkit, which is comprised of different modules that follow a general ethical strategy. The ethical strategy blueprint is discussed first, followed by a description of each of its phases and modules.

7.1 ETHICAL STRATEGY BLUEPRINT

A general overview of the ethical strategy blueprint is visualized in Figure 60. The strategy is composed by three phases, an Ethical Alignment phase, a subsequent phase that covers the Project Vision & Values, and a final stage that involves the introduction of a Moral Code for the project. These phases contain different modules to assist in the creation of a more ethical AI application. Each module is placed and organized on the phase where it would have the biggest impact if used, although, all the modules of the toolkit can be used separately if needed. The ethical toolkit is suited to be facilitated by any team that is working or will start an application that involves the usage of AI.

07



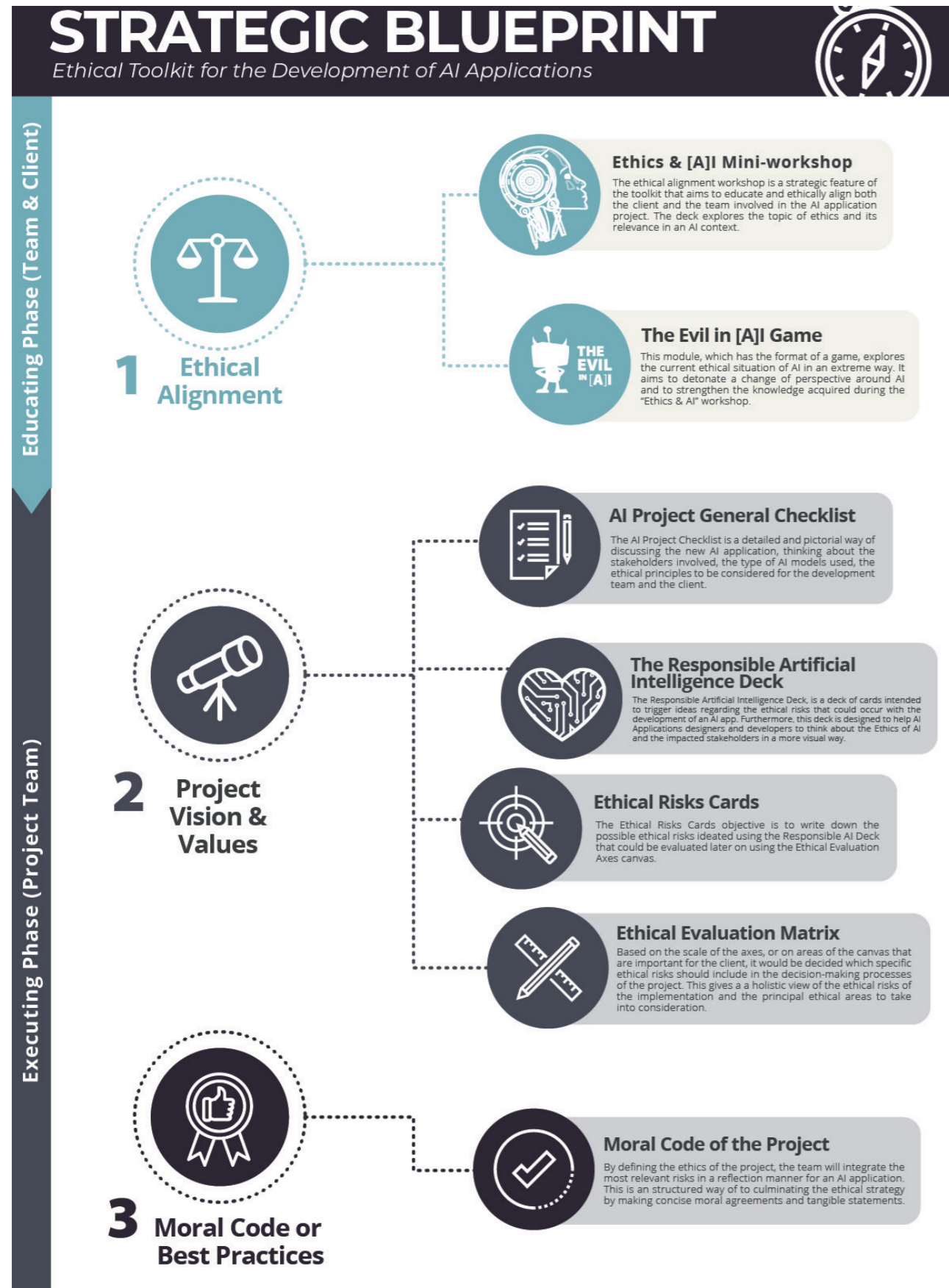


Figure 60: Overview of the Strategic Blueprint included in the Ethical Toolkit

7.2 ETHICAL ALIGNMENT PHASE

The first stage of the strategy was designed with the objective of educating and enhancing ethical alignment between the team and the client before formally starting an AI application project. From the interviewing stage it was observed that the stakeholders involved in the development of AI applications have a different definition of ethics or have an incomplete perspective of the discipline. This is somehow problematic as it was demonstrated with the prototypes that people tend to see ethics with an extremist view. This should not be the case as there are different ethical approaches towards dilemmas that depend on the context of the situation. This aligns with the idea that ethics is not about the right or wrong, but more about a gray area around the most preferred and the less preferred actions within a context. The modules of this phase are designed in order to strengthen this idea by using an introductory workshop to the topic of Ethics and AI. Furthermore, a gamified energizer is also proposed with the objective of creating awareness and triggering discussion around the current situation of the ethics of AI applications.

7.2.1 "Ethics & [A]I" Mini-workshop

Time: 1 - 1.5 hours in total preferably with the client in the initial stages of a project.

The ethical alignment workshop is a strategic feature of the toolkit that aims to educate and ethically align both the client and the team involved in the AI application project. An "ethics facilitator" should be chosen from the MOBGEN | Accenture Interactive team, who would present the workshop's deck to the company and the client. The deck explores the topic of ethics and its relevance in an AI context. The workshop begins with a world-view sensitizing exercise where the participants are asked to see themselves as an autonomous car and to pick up a destination. This exercise aims to spark imagination among the participants as well as creating an empathic idea of "human AI". Subsequently, the workshop continues with two examples of ethical

dilemmas, where the participants are asked to make an ethical decision and reflect upon their ethical views. Additionally, a definition of ethics is proposed based on the previous activities. In the final part of the workshop, three normative ethical approaches (consequentialism, deontological, virtue) are explained by involving the participants in different variations of the "trolley problem" using a "breakless" autonomous car. The workshop ends with the discussion of a couple of known real examples where development companies got affected due to unethical consequences of their AI. With the introduction of this workshop, it is expected that the participants understand the different ethical approaches that exist, as well as allowing them to identify their own ethical profile. Furthermore, the outcome will provide the participants an aligned and complete definition of ethics in an AI context. It is important to mention that this workshop is based on the "Machine Ethics Toolkit" work developed previously by Zhou (2018). The workshop package includes a deck of slides, a facilitator booklet, and extra support material.

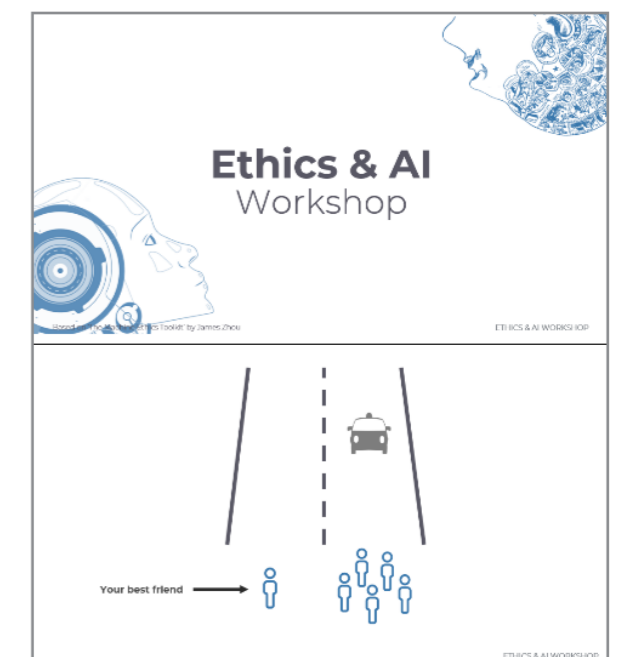


Figure 61: Samples of the Presentation Deck slides of the Ethics & [A]I mini workshop.

7.2.2 The Evil in [A]I Energizer

Time: From 30 minutes to 1 hour in total preferably with the client in the initial stages of a project.

This module, which has the format of a game, explores the current ethical situation of AI in an extreme way. It aims to detonate a change of perspective around AI, and to strengthen the knowledge acquired during the “Ethics & [A]I” workshop. The game consists on two decks of cards. The first deck contains “AI cards” with examples of current AI applications. The other deck contains “Evil cards” that expose unethical situations. First, participants are asked to randomly pick one card of the “AI cards” deck, which will be its AI system for the round. Then, each participant should randomly pick an in the same way, from the “Evil cards” deck. After this, all the players have to show and loudly read the combination they got (even if it is a non working combination). Afterwards, everybody

should vote about which participant has the most unethical example of all (from the combinations). After the discussion is done, the winner selected would gain one “Evil point”. The first participant to collect 4 Evil points is the winner. It is important to notice that all the card configurations should be written down on post-its after each round has passed. With this, after a few rounds, the team will have a list of unethical AI scenarios that would make the participants reflect upon unethical AI and think about the possible unintended future scenarios. One of the most important objectives of the game is to generate a discussion towards ethics in a playful manner, which would also help the participants set an aligned ethical view for the project. A first version of the game was validated with a couple of AI researchers from the TU Delft, which gave valuable feedback that was considered during the next iterations.



Figure 62: Overview of the Evil in [A]I game.

7.3 PROJECT VISION & VALUES

7.3.1 AI Project General Checklist

Time: 1-2 hours

The AI Project Checklist is a detailed and pictorial way of discussing the new AI application, a way of thinking about the stakeholders involved, the type of AI models used, and the ethical principles to be considered for the development team and the client. Both team and client will discuss the requirements and expectations of the AI application and write them down on post-its, so the content could be modified later. It is essential for the team to create a consensus around the canvas, to set the correct specifications with a proper argument behind in

order to have an appropriate overview of the project. The output of this exercise is a canvas filled with initial proposals, ideas, and ethical considerations of the AI application project. This canvas must be placed in a visible spot within the space selected for the development of the application. The team might use the Responsible AI Deck for triggering the ethical discussion. This module enhances the alignment during the initial decision making, establishes some ethical considerations to take into account at the beginning of the project, and generates awareness regarding the ethical impact in general.

AI APP PROJECT CHECKLIST

Name of the Project _____

1 Goal _____

2 Team Accountable for App
Client: _____ MOBGEN | Accenture Interactive

3 Context of the Project
e.g. AI application for healthcare

4 Stakeholders Impacted
Primary: _____ Unexpected/Unintended: _____

5 Type of Impact
Financial, Emotional, Access to goods, Property, Reputation, Life / Safety, Privacy, Liberty / Freedom, Rights / IP

6 AI Algorithm & Learning Model used _____

7 Data Type
Human Data, Non-Human Data

8 Data Source
GDPR compliant, Trustworthy

9 Ethical Principles Checklist (Order by relevance for project High(+) / Med(0) / Low(-))

..... Data Privacy Honest Communication Human Well-Being
..... Data Safety Accountability Governance
..... Explainability Fairness User Safety
..... Transparency Value Alignment	

Figure 63: Overview of the AI App Project Checklist

7.3.2 The Responsible Artificial Intelligence Deck

Time: 1-2 hours

The Responsible Artificial Intelligence Deck, is a deck of cards intended to trigger ideas regarding the ethical risks that could occur with the development of an AI app. Furthermore, this deck is designed to help AI Applications designers and developers to think about the Ethics of AI and the impacted stakeholders in a more visual way. It also features a “Risk/Opportunities” card, which is intended to trigger ideas regarding how to make an ethical principle an opportunity instead of only looking ethics as “compliance” in case something goes wrong. By selecting a specific card from the “AI Context” section (the most related to the intended AI application context), any development team could use some stakeholders and AI ethical principles cards to have a clear overview of the stakeholders involved and the ethical considerations that should be taken into account for the specific project. This relatable and playful manner to face the responsible development of AI is expected to prompt discussion and alignment among the members of the development team.

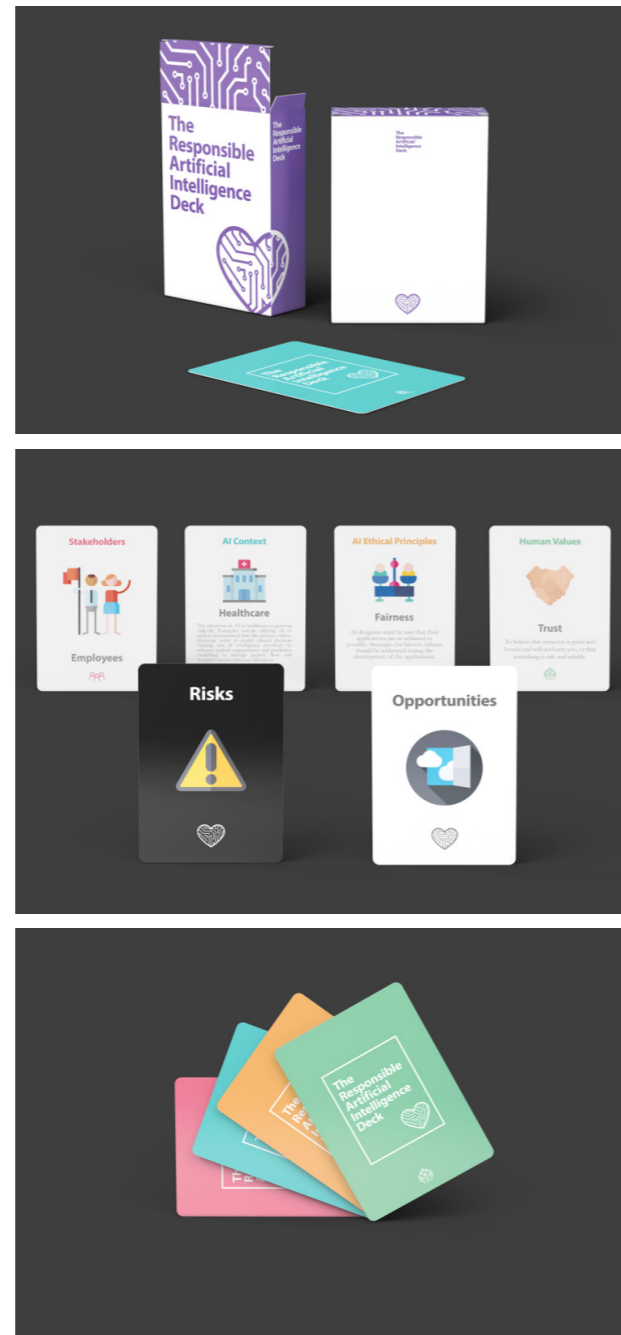


Figure 64: An overview of the final version of the Responsible Artificial Intelligence Deck

7.3.3 Ethical Evaluation Axes

Half a day

This exercise comprise of Ethical Risk Cards and an Ethical Evaluation Axes canvas. The former's objective is to write down the possible ethical risks that each participant ideated using the Responsible AI Deck. The idea of the use of the Ethical Risk Cards is to write down the biggest amount of ethical risks possible by each participants so these are not lost in the process and could be evaluated later on using the Ethical Evaluation Axes canvas. The evaluation canvas features a couple of axes where the “y” axis is for the level of “Impact” the risk could have for all the stakeholders involved (i.e. children, business people, businesses). The “x” axis is for the “Likelihood” of the risk occurring. The participants of the workshop should position the ethical risks cards on the “x” axis first. It is important to enhance some discussion regarding the reason of the positioning. After this was done, then the other axis should be discussed and positioned as well. Based on the scale of the axes, or on areas of the canvas that are important for the client, the team would decide which specific

ethical risks they should include in the decision-making processes of the project. With this exercise the team can get a holistic view of the ethical risks of the implementation and the principal ethical areas to take into consideration. It is important to mention that the evaluation axis can be referenced and modified throughout the project.

ETHICAL RISK CARD

AI Context

Stakeholder(s) Affected

AI Ethical Principle(s)

Risk Identified

Figure 65: Overview of the Ethical Risk Cards included in the Ethical Toolkit

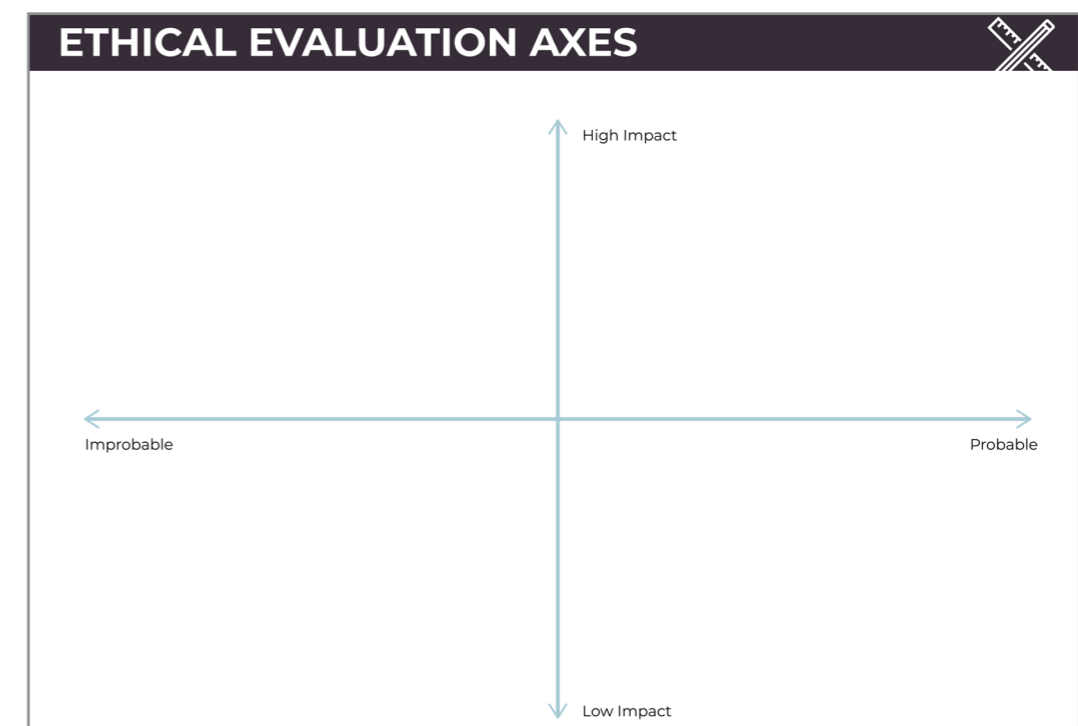


Figure 66: Overview of the Ethical Evaluation Axes included in the Ethical Toolkit

7.4 THE AI MORAL CODE





7.4.1 Project's Moral Code

Time: 1 hour

This module is intended to stimulate an ethical responsibility for the project. By defining the ethics of the project, the team will integrate the most relevant risks in a reflection manner for an AI application. This is an structured way of to culminating the ethical strategy by making concise moral agreements and tangible statements. The moral code establishes a scale of moral acceptance for the project. The team will define which actions are always or never adequate, as well as the middle gray area of what is acceptable and unacceptable if a condition is met. This is done by analyzing

the red areas of the Ethical Evaluation Axes and formulating rules that take into consideration these extremes. For instance, if one of the Ethical Risks discovered has to do with the safety of the users, then this could be added as a "moral rule" to be avoided in the future (i.e. It is never OK to attempt to the safety of users in any way). With this qualitative assessment, it is expected that a more ethically robust outcome is achieved. The output of this module is a moral agreement that could be translated to implementable product specifications for the AI application. It also intends to prompt discussion regarding changes that might involve ethical consequences later in the development process.

PROJECT'S MORAL CODE ✔

It's always OK to_____.	It's acceptable to_____ if_____	It's unacceptable to_____ if_____	It's never OK to_____.
			

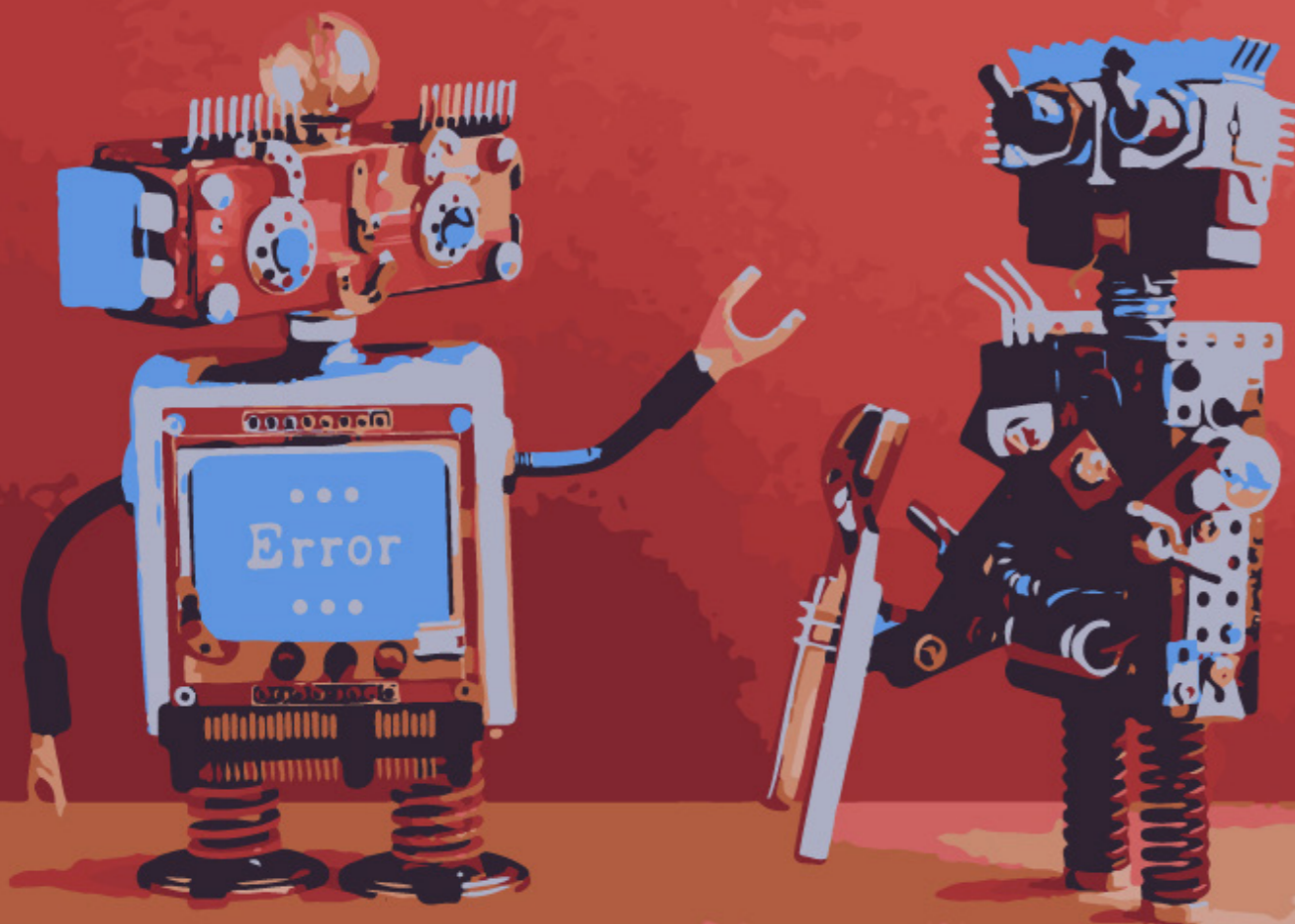
Notes:

Figure 67: Final version of the Project's Moral Code canvas.

7.5 CHAPTER CONCLUSIONS

This section describes the final design of the Ethical Toolkit strategy and modules. Six modules were presented and described including their inputs, processes, expected outcomes, and ethical approaches. The designed modules are intended to help with the creation of a more ethical AI application by stimulating awareness and ethical decisions in an original and iterative manner. The validation of the concepts with external stakeholders and internal stakeholders from MOBGEN | Accenture Interactive is discussed in the next section.

08



chapter 8

Design Validation of the Ethical Toolkit

This chapter provides an overview of the validations made for the Ethical Toolkit. Two validation setups were organized involving external stakeholders and internal stakeholders of MOBGEN | Accenture Interactive. The validation results will lead to the final recommendations discussed in the next chapter.

8.1 SET-UP OF VALIDATION

A validation stage was performed in order to evaluate the way and to what extent the proposed solution helps the stakeholders involved in the project. Special attention was taken toward the needs of the company as feedback was requested continuously in order to tailor the toolkit to their current circumstances.

The main objective of the validation stage was to evaluate the ethical toolkit according to the goals of the project, as well as the desirability, feasibility and viability of each of the workshop sections. This validation process involved several studies within MOBGEN | Accenture Interactive and also from external stakeholders like AI researchers, Ethics of Technology researchers, both from the TU Delft, and AI designers and developers from FJORD.

Several tests were made, for the external validation phase, involving the Responsible Artificial Intelligence Deck. These tests took place at the TU Delft EMMCS and TPM Faculties. Furthermore, a couple of session with professional AI designers, engineers, and data scientists were performed remotely from the offices of FJORD in Dublin, Ireland (Accenture's R&D Center). On the other

hand, for the internal validation, two test runs were made in different occasions for both parts of the strategic blueprint. The first part, which includes the Ethics of AI workshop and the Evil in [A]I game, was tested during a MOBGEN | Accenture Interactive experience chapter meeting. Furthermore, the second part of the blueprint was tested through a series of sessions made with a team from MOBGEN | Accenture Interactive.

There were evaluation discussions after all the validation sessions made during the project. Both, positive and negative feedback were collected from these discussions, which helped in the refining of the final design of the ethical solution. I personally provided instructions and facilitation aid during most of the validating sessions.

8.2 EXTERNAL STAKEHOLDER VALIDATION

This section elaborates on the validation of the ethical toolkit proposed with the support from some external stakeholders. The focus of the section relies on validating the desirability of the toolkit from an external point of view. The following subsections elaborate on the insights and feedback collected.

8.2.1 Validation session with AI researchers

A couple of validation sessions were performed with the assistance of two TU Delft researchers from the EEMCS Faculty. An Assistant Professor and a Postdoctoral Researcher from the Pattern Recognition and Bioinformatics department supported the sessions with their experience on building AI models from scratch. Firstly, the Evil in [AI] game was tested to find if the mechanics was adequate for a technical audience. Secondly, the Responsible Artificial Intelligence Deck was tested regarding the researcher's intentions of creating a company. The value of both ethical methods was clearly expressed as a way of ethical discussion and a tool for structuring relevant information for the stakeholders' projects.

Main insights from The Evil in [AI] validation

The response to the game was optimistic, however, the researchers found the first version of the game hard and confusing. This was mainly due to the complicated game mechanics regarding find an evil AI from the trigger cards.

Make it simple

The participants of the test recommended to make the game simpler. By making people think about too many different aspects of AI, the game becomes confusing.

Discussion Intention

The researchers expressed their positive feedback towards using the game as a mean of discussion towards bad consequences of AI. Similar insights regarding "the bad side of ethics" have been discovered in previous stages of the project.

Main insights from TRAIID validation

The participants found the responsible deck highly functional for their future spin-off intentions. The part they found most interesting was the Risks/Opportunities card, which allowed them to think about the ethical considerations as an added value for their product. Nevertheless, they also found that the ideation phase (of thinking about opportunities) was somehow complicated for them.



Figure 68: First external validation tests of the Responsible AI Deck done with the AI researchers.

"I personally find it difficult to think about a negative, or an evil AI, that covers all the requirements. Like that one, a marketing context for elders that is unfair...I consider marketing unethical in the first place"

Postdoctoral Researcher -EEMCS Faculty TU Delft

"I really like the game as a way to start a discussion about ethics. Mostly because of the visuals"

Assistant Professor - EEMCS Faculty TU Delft

Instruction Cards

Something that popped out during the test was the unfamiliarity that researchers have with this kind of design methods. They mentioned that including a set of instructions (or an instructions card) would be a good idea to help people becoming familiar with the different possible uses.

Visual Understanding

The participants appreciated the visuals of the cards which made them understand the Ethical Principles in a more concrete way, and overall, have a graphic representation of details that they might have overlooked at the beginning of a project.



Figure 69: First external validation tests of the Evil in [AI] game done alongside the AI researchers.

8.2.2 Validation of the Responsible AI Deck with FJORD-The Dock

A remote validation session was performed with the participation of people from FJORD - The Dock, which is Accenture's multidisciplinary research and incubation hub. "The Dock" is composed by a group of designers, researchers, and experts in artificial intelligence, advanced analytics, IoT, blockchain, security and mixed reality, that explore and experience how digital and emerging technologies will transform businesses and society. The remote session involved the participation of 8

"Does this includes a set of instructions? We would need to have a set of instructions to use it next time if you are not available."

Assistant Professor - EEMCS Faculty TU Delft

"This is really interesting and helpful for our spin-off. I think I really understand the methodology, like the steps of the methodology, but how to come up with actual, like opportunities and risks, that is still kind of problematic, but I like the methodology. It's like simple enough. And it gives you quite a lot of information that you can use. So that's why I asked, but, I think it's pretty pretty good."

Postdoctoral Researcher -EEMCS Faculty TU Delft

employees and it lasted approximately 1 hour. The participants background ranged from interaction design and service design, to software engineering. The Responsible AI Deck was tested during the session. After a short round of instruction where a general overview of the project was presented, two tests were performed by splitting the attendees into two groups. Firstly, each group grabbed 4 random cards (one per section) and used that configuration to trigger a conversation regarding the ethics of the situation. Then, the participants engaged in a discussion round with a "free style" use of the cards.

Both test rounds included an discussion session at the end where the participants provided some of their feedback about the deck and its advantages and disadvantages.

Main insights from TRAIID validation

The response to the Deck was mainly positive. Although, a few negative comments were collected, these are important since they reflect the current context of ethics within AI.

Conversation Mode: On

One of the most important insights gathered during the session was that the deck triggered a lot of conversation around the topic of ethics. All the different sections of the deck gave the participants a different perspective to think about their interactions with the technology. A participant mentioned that the deck could be useful to anticipate uncomfortable questions regarding the ethics of AI within projects.

"I think definitely just the way that that deck of cards can help facilitate conversation, that by itself is just a huge leap from where we are right now, where there's nothing to trigger that."

Systems & Service Design Lead - FJORD The Dock

"A lot of them have come up to me afterwards and said, that was really helpful and interesting. I want to do more stuff like that. I think you're like, you're kind of touching on a nerve. That's really interesting. Like people seem to really enjoy having the space to discuss this."

Systems & Service Design Lead - FJORD The Dock

"In my opinion, from what I've seen them go for it, they all work well together, like it makes as well, which I think is the biggest difficulty with with this is trying to get that it makes sense... And then, like for me, I'd love to do more with things and see as it seems to be multiple approaches you can take depending on the scenario. So it's which is kind of testament to the actual deck itself. I find it really useful."

Senior Interaction Designer - FJORD The Dock



Figure 70: Second external validation of the Responsible AI Deck with employees from FJORD - The Dock

Eye Opener

Some participants mentioned that the session was an eye opener for them since they had not considered the possibilities that the deck illustrated. For instance, one group discussed about AI powered marketing platforms card and the stakeholder cards containing children/teenagers, detonating an interesting conversation focused on participants that were parents.

Clients might want it

Both groups agreed that the cards were a huge advancement in the topic as there are currently little options that cover the responsible development of AI. They would use it while being with a client and definitely use it on a project.

Make it "Good vs. Evil"

A group of participants proposed an additional approach for the deck, to include an evil/good card. The idea is that the participant group is divided into two 2 teams, with each team taking a side during the discussion. According to the participants the answer for the ethical considerations should be in the resulting arguments ("in the middle of good and bad"). This is something that it is expected to be covered by the Evil in [A]I game, which was not introduced to this team.



Figure 71: Software engineers and service designers from FJORD - The Dock validating the Responsible AI Deck

"...by being transparent you actually introduced me to the fact that this stuff is doing can be a little bit dangerous, but like I never never even thought about it that way at all."

Senior Interaction Designer - FJORD The Dock

"This would be useful to bring into a client session, especially some suggestion of like stacking the deck with clients specific use cases. Yeah, exactly. You are talking to someone in the public sector, you could stock it with things that are relevant there. Have them then discuss in general, like ethics in that area."

Systems & Service Design Lead - FJORD The Dock

"Have you considered like...including an evil/good card? The way I see it, it would be pretty interesting to have two teams that discuss bad and good outcomes of the ethical discussion.... The answer about ethics should be in the middle ground .Yeah, that might be helpful, kinda like a 'red vs blue' exercise..."

Senior Interaction Designer - FJORD The Dock



Figure 72: UX designers and business developers from FJORD - The Dock validating the Responsible AI Deck

Aligning definitions

Something important that was mentioned during the feedback stage of the session was the different understanding that participants had about the ethical principles cards. For instance, the differences between “Data Safety” and “Data Privacy” were obvious for one group but hard to differentiate for the other. An alignment session, or just a sensitizing exercise, was proposed by the participants to avoid this situation.

Just a design tool

An interesting finding was presented by a participant with an engineering background. She mentioned that the tool was not useful to her daily job as it is more a “tool for designers”. This triggered a conversation with another designer in the session which followed after the validation session as the technical participant insisted in the limited role that engineers have during the creation of AI products, and how their opinion about ethics “might not be necessary”. This conversation between participants shows how different stakeholders see their role within the creation of AI applications. It also shows an urgent need of inclusive ethical methods for the creation of AI apps.

8.2.3 Validating toolkit with “Ethics of Technology” researcher

A quick validation session was executed with the assistance of an ethics of technology expert from the TPM Faculty of the TU Delft. Firstly, the Evil in [A]I game was tested to validate if the game would be a good choice for ethical education and discussion from an academic point of view. The Responsible Artificial Intelligence Deck was tested afterwards with the same objective.

Main insights from The Evil in [A]I validation

The response to the game by the ethics expert was mainly positive. Some constructive comments involved a way to measure the “ethical aspects” of the game in order to make it more logical to play for participants.



Figure 73: Third external validation of the Evil in [A]I game.

“I think what you tried to do here is fine, I mean... what you did is to create consciousness around the ethical danger of these situations...people will tend to intuitively measure what is right or wrong. Some people would even try to avoid the discussion with phrases like “but that depends”, however, the objective of the game is achieved because it is not intended to create an universal law of behavior, but only to discuss and educate people about ethics.”

Postdoctoral Researcher - TPM Faculty TU Delft

“You can also make a way of measuring winners and losers, like other board games...kinda like an ‘ethical value’...or make people vote who’s got the most unethical combination of all.”

Postdoctoral Researcher - TPM Faculty TU Delft

Main insights from TRAIID validation

The response to the Deck by the ethics expert was positive, however, he expressed a series of concerns regarding a possible “justification” for unethical actions. The comment was triggered due to the risk/opportunities card, which according to the expert’s opinion should be defined as an “opportunity to be more ethical” instead of making a “business opportunity out of ethical behaviour”. All the comments were considered for the further iterations of the Deck.



Figure 74: Third external validation of the Responsible AI Deck

“I think that the idea of discussing ethics with these cards is interesting, although, I am concerned about the ‘Opportunities’ card...it should be clearly defined as an opportunity to be ethical, not to use ethics to avoid responsibility.”

Postdoctoral Researcher - TPM Faculty TU Delft

“As an educative effort it is really nice, however, you must take into consideration that your tools are not used as a ‘justification’ mean for ‘doubtful ethics’ from the client side.”

Postdoctoral Researcher - TPM Faculty TU Delft

8.2.4 Discussion

It is noticeable that from an external stakeholder perspective, the toolkit created is useful as a conversation starter. Moreover, its educational purpose opens up the discussion around the design and development of AI applications. The visual composition of the cards proven to be one of its biggest strengths as it provided of details that would have been ignored otherwise. This is understandable since, as of today, there is not other example of a graphical ethical toolkit that has been published. Something that is important to mention is that the Ethical toolkit is considered more like a design tool made for support critical innovation/reflection rather than to a technical tool that helps to envisage solutions. This reflects the intention of the project, since it was discovered that within MOBGEN | Accenture interactive there is no ethical assessment of projects and the topic of ethics of AI is not in the top of people’s mind. Hence, it was important to ideate a proper solution that allows the company to establish themselves as leaders in the topic among other implementors.

8.3 INTERNAL STAKEHOLDER VALIDATION

This section elaborates on the validation of the ethical toolkit within MOBGEN | Accenture Interactive. The focus of the section relies on validating the desirability, feasibility, and viability of the toolkit proposed regarding the company context.

8.3.1 Validation of Ethical Alignment Stage within MOBGEN | Accenture Interactive

I had the opportunity to validate the first stage of the ethical strategy (ethical alignment) during an experience chapter meeting which is organized within the design community of MOBGEN | Accenture Interactive. Almost 30 participants attended the chapter and validation test which lasted around 40 minutes. The majority of the participants have a design background, however, there were creative technologists involved during the session, as well as business analysts and account managers from Accenture Interactive. The “Ethics & AI” workshop was tested first, followed by a test run of the Evil in [AI] game.



Figure 75: Internal validation test of the Ethics & [AI] mini-workshop performed during a MOBGEN | Accenture Interactive “experience chapter” meeting.

Main insights from the “Ethics & [AI]” workshop validation

The overall response to the workshop was mainly positive. Participants got engaged with the ethical dilemmas exercises and how it helps them thinking regarding ethics in general. The variations of the trolley problem gave the attendees a nice perspective of the different ethical approaches by using the context of AI as an example (breakless autonomous car).

Clear examples

One of the participants expressed his confusion regarding the different ethical approaches, mainly between a deontological and consequentialist approach. This comment sparked a conversation which was complemented by an example. This event show the need to provide the facilitator with more practical examples on how to explain the differences between ethical approaches.

“It is cool to learn more about the ethical approaches that exist, also the game is kinda entertaining. So this might create awareness with our clients.”

CCO - MOBGEN | Accenture Interactive

“I like the interactive element of the presentation but I wish maybe we went a bit deeper into the process of injecting ethics into the ideation process”

Service & Interaction Designer - MOBGEN | Accenture Interactive

“It wasn’t that clear to me, that’s why I asked about the differences between the ethical approaches.”

Creative Technologist - MOBGEN | Accenture Interactive

Main insights from the Evil in [AI] game validation

The response to the game was in general positive and entertaining for the participants. On the other hand, the game triggered a lot of confusion mainly due to the language used in the cards and the objective of the exercise. This might have been caused by the poor understanding of the topic in hand from some of the attendees. Ethics is a complex topic, mostly in a non-academic environment. Moreover, only a small part of the attending audience had a considerable knowledge of AI and its applications.

The main insights of this validation were:

- The game should contain an instructions booklet or board with graphical instructions to avoid the participants’ confusion.
- More tests within MOBGEN | Accenture Interactive and FJORD could expose more variations of the game with the same objective of discussing and reflecting upon the “Evil side of AI”.

8.3.2 Case Study with MOBGEN | Accenture Interactive team

A short validation session was performed with a development team of MOBGEN | Accenture Interactive. The project in development involved the creation of an AI-powered real estate online platform, which would make use of a smart algorithm to inform investors about good opportunities in the market. Three designers participated during the session that lasted 1 hour, where The Responsible AI Deck was tested. The value of the Deck was clearly expressed as a way of searching for ethical opportunities and risks in the project.

Main insights from TRAIID validation

The response to the Deck was mainly positive. The designers were able to prompt interesting ideas with the assistance of the AI Ethical Principles



Figure 76: Internal validation test of the Evil in [AI] game done during a MOBGEN | Accenture Interactive “experience chapter” meeting.

“This would be a great addition to our case because yeah, it actually triggers some good ideas, mostly opportunities, which is interesting.”

Creative Technologist - MOBGEN | Accenture Interactive

“I think maybe if there was a bit of structure it could be even more relevant...I would feel a bit lost if you weren’t here explaining how to use them”

Creative Technologist - MOBGEN | Accenture Interactive



Figure 77: Internal validation test of the Responsible AI Deck with an AI App development team from MOBGEN | Accenture Interactive

cards. Furthermore, by using the deck to make an analysis of the project, the team was able to identify some assumptions they had in a more visual and appealing way. The deck also helped the designers to come up with opportunities backed up with the ethical principles proposed. For example, the real estate industry relies on good investments made from valuable properties that are not interesting at first sight, which in other terms it could be an outlier from a biased set of properties. This intentional bias was identified while using the deck and it represents a case where an ethical consideration could be used as an opportunity. An important piece of feedback collected from the validation session was to include the real estate into the AI Context section and, as previously requested, include a set of instructions to use the Deck.

8.3.3 Complete Ethical Toolkit Validation on client case study

A final validation session, which included the testing of all the modules of the toolkit in a workshop manner, was performed with the participation of a development team from MOBGEN | Accenture Interactive. The session involved the participation of 5 employees and it lasted approximately 2 hours. The participants role in the organization ranged from service designers to creative technologist and software developers. The whole ethical strategy was tested during the session as the participants were involved in a project where the implementation of AI is expected in a future stage. Unfortunately, there were no representative parties from the client of the project, nevertheless, some interesting insights about the way employees could use the developed solutions in this and future projects were discussed and collected.

The session began with a short round of instruction where a general overview of the project was presented. The validation run included two “discussion and feedback” sessions which were conducted at the end of each stage of the ethical strategy. A summary of the main findings is explored next:

Main insights from the “Ethical Alignment” Stage (#1)

The overall response to the first part of the ethical strategy was mainly positive, although, a few constructive comments were mentioned regarding the tested version of “Evil in [AI]” game and how it could change to have a bigger impact on current MOBGEN | Accenture Interactive clients. It is important to mention that a different game version from the one proposed in Chapter 7 was tested during this session, where the cards (and a dice) were used in a “board game” manner due to the limited time span available. On the other hand, regarding the “Ethics & [AI]” mini-workshop, a few comments were collected mostly related to the complexity of the ethical frameworks. It was recommended to use a printed takeaway card with the definitions written

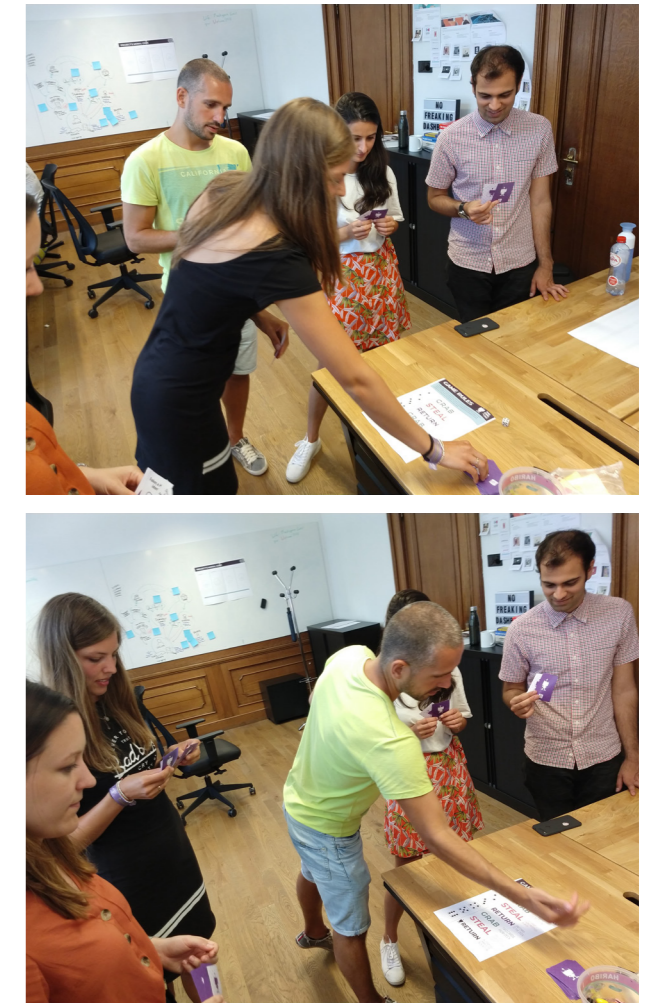


Figure 78: Test run validation of the “Evil in [AI]” game.

“It would be nice to look more into the content rather than into a gameplay, because right now I feel like we are only counting numbers and the cards are really beautifully designed...it is like a missed opportunity”

Service Designer - MOBGEN | Accenture Interactive

“Maybe try to come up with a convincing story on the winning combinations, like how evil is this....I would definitely use the cards during my facilitation sessions to spark ethical discussion, but without the game”

Product & Service Designer at MOBGEN | Accenture Interactive



Figure 79: Filling in the AI Project Checklist with the help of the validation session participants.



Figure 80: Final validation test of the Responsible AI Deck.

down and some examples (i.e. Trolley Problem) to promote a more solid understanding of the subject.

More discussion, less “board game”

As mentioned previously, participants recalled that the game is indeed fun but the tested version (game with dice) would become only a “board game”, which do not allows time for discussion. More importantly, the participants felt this was a waste of potential from the cards which could be used for discussing about ethics of AI and share the knowledge and different ideas that the “Ethics & [A] I” mini-workshop triggered. This recommendations were listed and implemented for the final version of the ethical toolkit, where the game is based on the discussion regarding the most “unethical” combination of cards.

Main insights from the “Project & Vision Values” Stage (#2)

The response to the modules of the second stage was optimal, however, it triggered a compelling discussion about the topic of ethics and the assessment of AI projects. Some modules got a nice acceptance from the participants, for example, the AI Project Checklist was filled in in a seamless way. It also helped them to discuss and discover some things that might have been overlooked like the type of impact the AI system would produce and the type of data that would be used. Moreover, some suggestions were made regarding the content of the canvas like the differences between “Human and non-human data”.

Are we accountable?

A curious scene that happened during the AI Project Checklist filling of the “Accountability section” was the perception of unaccountability that most of the participants of the team had regarding the project. As it has been previously documented (Van de Poel, 2011), a sense of passive responsibility is commonly present among designers and engineers. It was observed during this exercise that this might also be the case with the team due to the role of “external consultants” that MOBGEN | Accenture

“Wait, are we all responsible? (...) It has been documented that if you assign the responsibility to the whole team, then nothing would be done, contrary to give the whole responsibility to one person, a manager or so...this way productivity might rise”

Service Designer - MOBGEN | Accenture Interactive

“I kinda struggle to think about how we act towards the stakeholders, as well as the client.... because some stakeholders are more important for us than for them.”

Product & Service Designer at MOBGEN | Accenture Interactive

Interactive employees play within the project. Some suggestions about why this happened among the development team were shared as well by some of the participants as they believe that the feeling of active responsibility from the group might affect the productivity of the team as a whole. Because of this, they believe that only the client was responsible as well as a single “manager” or accountable person from the company. From an ethical point of view, the way this was addressed might affect the perception of ethics in the future of the project as this propiciates the notion of superficiality about ethics.

Involving the client would be good (And some AI experts as well)

Some participants mentioned the importance of including the client in the discussion as it would make things more clear and it would trigger alignment among them and the client. They also mentioned that without the correct guidance from experts in the topic of AI it would be difficult to fill the AI Project Checklist.

Do you have any example?

One important insight that is worth to mention is the difficulty that participants experienced during the identification of ethical risks. It was noticeable that their struggle had to do with the need of inspiration from examples as they were not familiar with the topic. Although, the deck provided a good visual overview of the ethical principles assessment,

the stakeholders involved and the context related, it does not take into consideration any example to get inspired from. This is something that was taken into consideration for the final design.

Main insights from the “Moral Code or Best Practices” Stage (#3)

It was noticeable that the participants were confused about the transition between the second and the third stages. As the complete toolkit was not tested before, it was never noticed that the final section of the workshop is quite complicated to execute. In the end, the “most moral choices” were based mostly on the ethical risks discovered on the second stage.



Figure 81: The results of the final validation test of the Risk Cards and Evaluation Axes.

Other General insights from the Validation

Session

At the end of the test session of the final section of the ethical strategy, there was a discussion about the reason behind the project and how it would be applied within the organization as an added value proposition. Some of the participants share their opinion about the main objective of the workshop and the difficulty, according to their experience, to sell this to their clients.

Another interesting suggestion came from one of the participants which played with the idea of including an internal facilitator that acts as an “expert in ethics” within the company. This would avoid some of the “stuck moments” that were experienced during the session as this expert would ensure a continuing line of thinking and support the participants during the ethical assessment process. This has already been proposed by Simons (April, 2019) which, among other things, would help to avoid a potential client to feel lost throughout the process.

Finally, an advice about including a section where AI could be explained in a fast way was mentioned by most of the participants. Although a couple of participants were familiar with the concept of AI and its ethics, the rest of the team were not and they had a rough time understanding some of the concepts during the test. This is an important insight since it has been taken for granted that the people which participated in the ethical strategy workshop already are familiar with AI and its related concepts.

“It would be better to think more about KPIs instead of ethical principles...this way it would be more appealing to our clients.”

Product & Service Designer - MOBGEN | Accenture Interactive

“I think that you would require somebody that is kind of an expert on the topic of ethics that pushes the facilitation of the workshop forward... there were several times of silence during the session, you should consider continuing the line of thinking to avoid clients feel lost on the process”

Creative Technologist - MOBGEN | Accenture Interactive

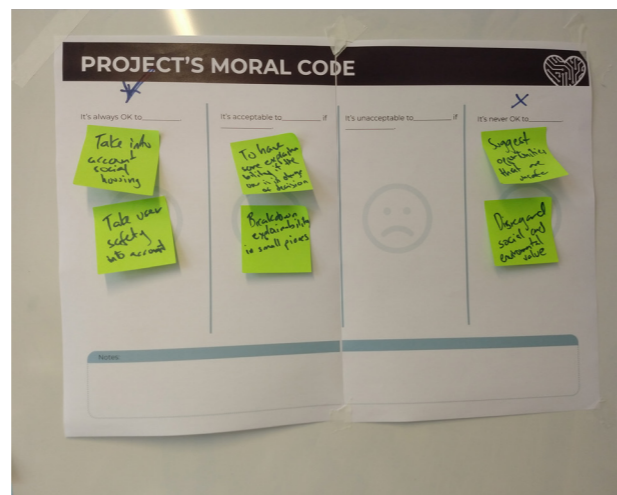


Figure 82: Overview of the Strategic Blueprint included in the Ethical Toolkit

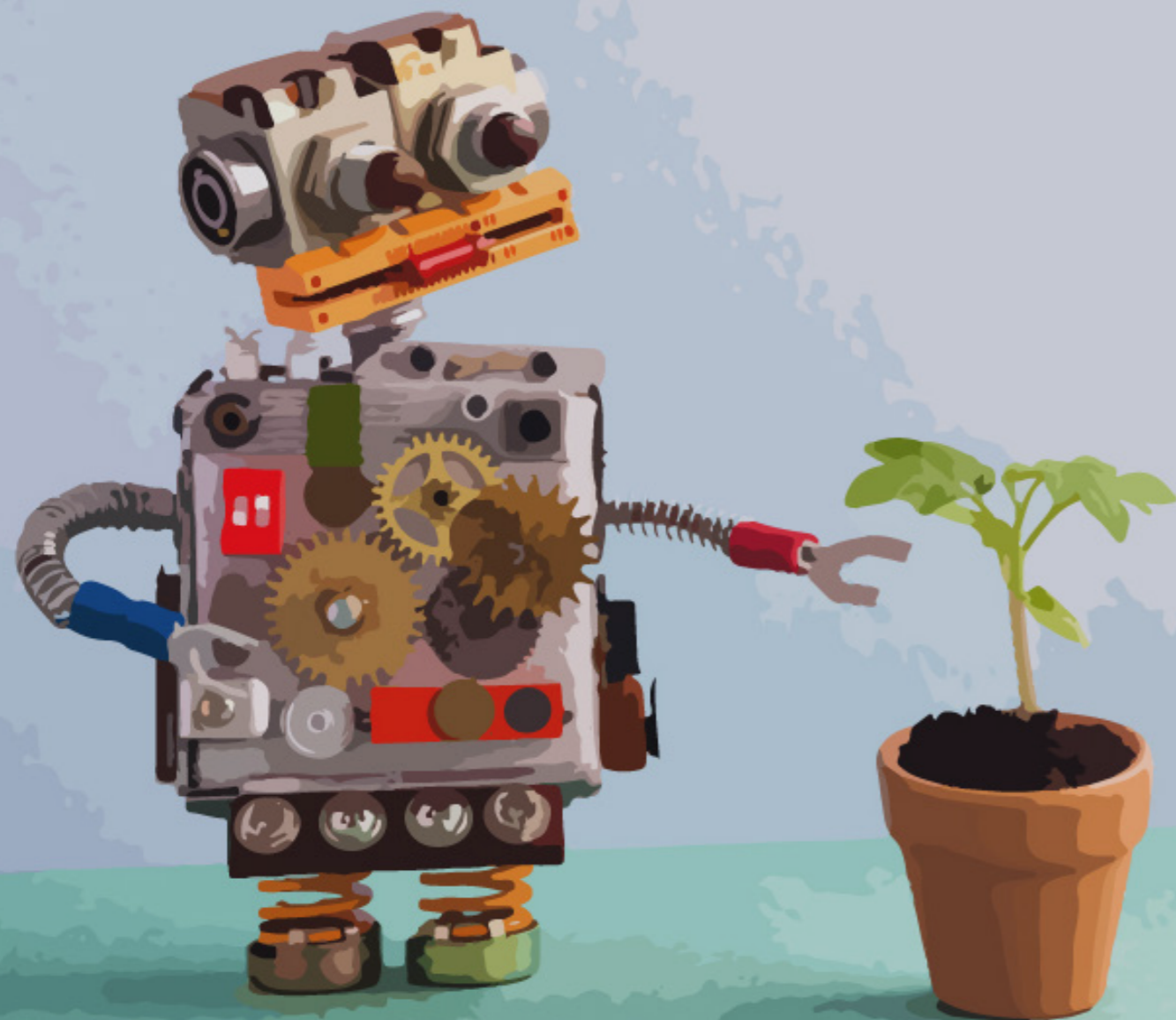
8.3.4 Discussion

This section shows the series of internal validation sessions that were performed with employees of MOBGEN | Accenture Interactive. It was noticed that from an internal stakeholder perspective, the toolkit is a useful asset for designers and developers, although it might require more work to adapt it completely to the facilitation processes of the organization. This is understandable to a certain extent as, due to the poor availability of AI projects within the company, it was difficult to organize full validation rounds. On the other hand, it was concluded that the modules work extremely well independently, as observed during the individual tests. This increases the value of the complete toolkit as an open source tool, since it could be adapted to other processes, as well for the use of other dependencies from Accenture, like FJORD or Storm Digital.

8.4 CHAPTER CONCLUSIONS

Several internal and external validation runs were conducted at the end of the project as described in this chapter. These sessions were intended for the collection of integral feedback and tips regarding the modules that compose the ethical toolkit. During the external validation sessions, several stakeholders were consulted like AI researchers, Ethics in Technology researchers, and a group of employees from FJORD - The Dock. One of the most important insights from this stage is the usefulness of the toolkit as a conversation starter around the ethics of AI. Its educational purpose opens up the discussion and propiciates critical thinking as well as ethical understanding of the problem. However, the toolkit does not trigger possible solutions to the associated ethical risks. Nevertheless, this could be an interesting topic for future research in the field of design-oriented ethical tools for AI. Likewise, during the internal validation sessions, it was noticed that although the toolkit was perceived as useful, it might require more work to adapt it completely to MOBGEN | Accenture Interactive facilitation processes. Most of the validation sessions were focused on testing the ethical toolkit modules individually. Thanks to this, it was observed that the modules work extremely well independently, which increases the value of the ethical toolkit within Accenture and its dependencies due to its customization potential.

09



chapter 9

Recommendations, Limitations & Implementation

This chapter explores the recommendations for MOBGEN | Accenture Interactive, an overview of the limitations of this project, and it also covers steps and other considerations required for the implementation of the ethical toolkit.

9.1 DISCUSSION & IMPLEMENTATION RECOMMENDATIONS

This section clarifies the implications and recommendations related to the implementation of the ethical toolkit and the theoretical general discussion of the complete project. The discussion of the project follows a detailed explanation of the research questions proposed.

9.1.1 General Discussion

Q1: What are the ethical views and future concerns of all the stakeholders involved in the development of AI applications?

Q1.2: How these insights could help to produce an Ethical Framework for AI?

The general ethical views and final concerns of the stakeholders involved in the development of AI applications were described by means of a series of semi-structured interviews, a creative discussion, and a prototype presentation. During this stage it was found that a lack of knowledge and integration of ethics in current processes is present in the development of AI applications. This is something that has been discussed continuously in previous research. For example, Although, there is a noticeable interest and awareness from most of the interviewed stakeholders, a feeling of “passive

responsibility” is still present. Furthermore, the results obtained during the interviews reflected a notion that ethics is still perceived as a superficial topic. This might have been provoked by the absence of tools to ethically assess the outcome of AI projects, which is something that was also found as an important insight.

The results obtained in the empirical stage enhanced the idea of the need of an Ethical Framework that could be used in a more intuitive and visual way and that covers up most of the topics studied. Certainly, this was accomplished with the further construction of the framework by the application of several design methods and some analysis sessions performed personally and with the assistance of TU Delft students and MOBGEN | Accenture Interactive employees.

Q2: How to support MOBGEN | Accenture Interactive with tools for a responsible development of AI applications?

It was discovered during this project that the best way to support MOBGEN | Accenture Interactive is in the form of a workshop that could cover the topic of AI ethics. This was decided after an extensive internal research effort with semi-structured interviews and generative exercises that were executed within the organization. The insights collected from the interviews included the different processes that are followed within the company at different levels. The research stage sparked interest also from upper management and it started a conversation around how to implement the solution to have a strategic fit in the company's processes. Moreover, due to the different levels of hierarchy in the organization and the different types of contact with the client, the ethical solution would have to cover both high and low strategic levels in order to have a robust ethical solution.

A robust strategy composed by two phases was develop throughout the project, and it aims to provide an introductory "educational" phase followed by a more practical one focused on the execution of the project. This configuration was selected due to the insights gathered on the way ethics is perceived within the company, and also by its clients. The strategy aims to trigger and support dialogue about the ethics of AI and to motivate a critical innovation culture within MOBGEN | Accenture Interactive.

Q3: What are the strategies that MOBGEN | Accenture Interactive could follow to provide an added value proposition to their customers through the ethical uptake of AI applications?

As mentioned previously, the internal research phase triggered ideas about how to create an added value with the ethical strategy conceived in the form of a toolkit. Some of the ideas were to include the modules generated into the mobile application

made for workshop facilitation called ENSO, which was developed by MOBGEN | Accenture Interactive. Moreover, it was also concluded that a way of introducing this to clients would be as part of the design sprints that the company does with its clients in a "full day" ethical workshop fashion. It is important to mention that the ethical toolkit was arranged in a way to fulfill this last statement. In this way, the ethical workshop would be divided in two phases and would try to bring value to both MOBGEN | Accenture Interactive and its clients.

9.1.2 Implications of a correct implementation

In order to have a beneficial fit and commercial advantage of the tool, several implications must be considered by MOBGEN | Accenture Interactive. By doing this, the company will ensure that a strategic alignment towards ethics would give them an added value for their clients, which might bring profitable benefits in the long run. The implications to take into consideration for a correct implementation of the ethical toolkit are arranged and discussed per module.



01 Ethical Strategy Blueprint

It is recommended that the ethical strategy blueprint keeps its nature of an "non-finished" solution so any employee from MOBGEN | Accenture Interactive could modify it in the future. This would foster a more ethical and diverse ethical strategy. In order to spread the solution among all the Accenture dependencies, the digitalization of the modules could be performed and promoted within the internal communication channels. One solution would also be to add the information generated in the modules into the mobile application made for workshop facilitation called ENSO, which was developed by MOBGEN | Accenture Interactive. Furthermore, the ethical strategy could also be added to Accenture's global digital platform "_FORM_", where a vast amount of design methodologies are shared as well. The use of each module individually is strongly encouraged as it would be helpful for the company to personalize the content depending on the nature of the AI application they are working on. I am currently in contact with a team of FJORD - The Dock to further develop a digital version of the content and to include it into their processes.



02 Ethics & [AI] Mini-workshop & Evil in [AI] game

It is strongly recommended to promote the first stage of the ethical strategy (mini workshop and game) when conversations with a client begin about including AI into an application. This is something that was found at the beginning of the project and with the multiple talks that were made with experts in ethics. A stage of ethical education is required first to remind people about the importance of the topic and the different ethical views that exist.

One of the most interesting insights collected was the need of more common words to explain the ethical jargon to people during the mini-workshop as well as the use of practical examples to explain the differences between ethical approaches. In the case of the game, several changes were made after the different series of validation performed. The final version makes the participants get familiar with the topic in a extremely simple and relatable manner with the inclusion of a voting system that eventually would lead to ethical discussion and the assessment of the winner. It is important to mention that the game could be also modified as MOBGEN | Accenture Interactive should maintain its "open source" status. Additionally, and as suggested during the validation phase, it is recommended to give the client a previous workshop where the basic concepts of AI are discussed. This way, all the participants of the ethical strategy workshop would be familiar with the context of AI.



03 AI Project Checklist

Perhaps one of the most useful parts of the toolkit, the project checklist is a straightforward compilation of the relevant information of the project. Because of this, it is recommended that it is kept in a visible space throughout the duration of the AI app project. Another recommendation that also was addressed during the validation phase is to include the client always during the filling in of the canvas. In this way, both the developing team and the client would get involved into an ethical alignment phase that begins with writing their names in the "Accountability" form (active responsibility) and it ends up by deciding the order of importance of the ethical principles for the project based on the client's opinion. Additionally, it is recommended to include a limit when talking about the direct and indirect (unintended) stakeholders of the project to avoid

falling into a generalization of stakeholders. The specification of these, would enhance a more ethical overview of the people affected, which is something that is often overlooked.



04 The Responsible AI Deck & The Ethical Evaluation Axes

As mentioned before in the validation section, it is recommended to assign a facilitator that is familiarized with ethics and the ethical risks of AI (it can also be trained in the topic as well). This should not be confused with the creation of an “ethical coach” as proposed in previous research. The recommended facilitator would have the responsibility of helping the continuation of ethical thinking throughout the workshop, but it is not accountable for encouraging employees from MOBGEN | Accenture Interactive to be more ethical. This is unrealistic, unfeasible, and is generally addressed by other dependencies within organizations, in the case of this project, the Compliance Department of Accenture. Additionally, the open source state of the toolkit suggests that it can be modified and adapted specifically to fulfill certain design sprint objectives. This is something encouraged and recommended as the Responsible AI Deck is not future proof. Moreover, due to the exponential interest and progressive development of the field AI ethics, a continuous improvement and actualization of the deck is encouraged. This way, the deck would remain latent for a longer period of time as it would include the new developments around AI.



05 The Project's Moral Code

In the case of the moral code of the project it is recommended to look for ways to translate the “ethical specifications” into actual implementable solutions or “technical specifications”. Due to the time frame of the project, this was one of the parts that could not be researched further. This was justified by the fact that there were no previous solutions on the topic of ethics of AI within the company, so the ethical toolkit would work as a starting point on the topic between MOBGEN | Accenture Interactive and their clients.

9.2 RESEARCH LIMITATIONS & FUTURE RESEARCH

This section elaborates on the limitations of the current research project and on future topics for further research in the field of ethics of AI.

9.2.1 Research limitations

The objective of this thesis is to support MOBGEN | Accenture Interactive with the ethical assessment of their application projects that involve the implementation of any type of AI. This is made with the intention of provide an added value proposition for their clients. An extensive research effort was made to reach to this point which includes the current state of the field of AI ethics, several empirical interventions and the development of a theoretical framework. On the other hand, the current research effort have some limitations that is worth it to discuss.

01 Limitations of the first stage of the project (First Diamond)

In the first stage of the project, a literature study of AI, Ethics and the combination of both was executed in order to generate a robust theoretical ground. Furthermore, is not much to say that these fields are vast and complex on their own, as the amount of information about the topics is unbearable for a 20 week project. Because of this, it is expected that the identification of some insights were missed out. It is important to mention that a few assumptions were made based on the information found, which were validated to a certain extent, during the empirical research phase.

Additionally, on the subsequent empirical stages, several assumptions were made and a general approach was taken on the ethical role of the stakeholders involved in the development of AI applications. This means that instead of focusing on a single group of stakeholders, which would belong to Accenture or any of its dependencies, the study focuses on the general stakeholders that would be impacted with the responsible development of an AI application throughout the value chain of

the development (AI Researchers, Implementors, Clients, Ethicists, and Users). Although, it is included in the stakeholder map presented in Chapter 5, it was decided to let the Government out of the scope of research due to the difficulty of getting in contact with the correct dependencies. The idea behind selecting this group of stakeholders was of generating a robust ethical tool conceived in a co-creative/crowdsourced manner. However, it is recommended to conduct further research in the ethical impact of AI applications in the whole spectrum of the value chain, as well as from the perspective of each stakeholder. Something that has already been encouraged by Simons (April, 2019).

Certainly, it is the first time that the author make use of action research cycles in the development of a project. Hence, a lot of work still needs to be done in order to master this research method to make the best out it in future projects. The project that is described in this report has been developed in an iterative way and, most of the time, in a simultaneous manner. Because of this, it is important to mention that the process followed in this thesis is not strictly chronological as some results from one stage impacted in the same stage as well as in future stages.

02 Limitations of the second stage of the project (Second Diamond)

The framework proposed in this project is founded in the literature review executed, the continuous meetings with experts in the topic, and the diverse empirical studies performed. Moreover, the final design of the ethical toolkit is based on this framework. Although several validation sessions for the framework were performed, there are still some details that can be improved and that were

not optimized due to the time frame of this project. For example, the Evil in [A]I game has a lot of potential for future iterations where different game mechanics could be envisioned depending on the ethical message that is required. Moreover, a module or section of the workshop where the participants can study the Responsible AI Deck in order to align the definitions could be added afterwards. This is related to the comments obtained during the external validations made with FJORD, where the terms “Data Privacy” and “Data Safety” were considered similar by some participants. In addition to this, some trade-offs were made regarding the detail to which the solutions were developed, again due to the lack of time for developing the final solution.

Certainly, for the case of the validation sessions, still some work needs to be done as more tests should be performed with actual clients. The author has been currently in contact with people from FJORD - The Dock that is deeply interested in working further with the developed toolkit and perform the required validations. Due to the time frame for the project, and the availability of the organization, only a single complete test could be performed. It is important to mention that each section of the toolkit was validated several times independently with good results. Nevertheless, the impact from the complete toolkit has not been researched in depth within a project from MOBGEN | Accenture Interactive or with the participation of a client. This is because there were no current projects within the organization that involved the use of AI. A next step here would be that the company test the entire ethical strategy in a whole day workshop session for a real client, and would be able to measure the actual impact of the solution. This could be done by any facilitator (i.e. product & service designers) from MOBGEN | Accenture Interactive by introducing the activities proposed in the ethical strategic blueprint on a “design sprint” process where a project deals with the implementation of AI. The deliverables attached to this thesis contain an instructions booklet to guide the facilitator throughout the

ethical toolkit. Nevertheless, it is recommended to the facilitators to freely customize the use of the tools depending on their own personal facilitation style and the type of project/client.

9.2.2 Future Research

This project resulted in an ethical toolkit that could be applied for the development of AI applications, however, the results obtained could generate the first steps towards further research in the topic.

01 Ethical frameworks developed using design methods

It might be valuable to start using design methods in the development of ethical frameworks for AI. Most of the frameworks available do not foster a visual and intuitive way of either the perception of ethics or the assessment of the ethical impact.

02 Measuring ethical impact

Measuring the ethical impact of AI systems would be an interesting topic for further research due to the lack of current solutions towards the topic.

03 Research on the ethics of AI from a design perspective

Designers have a unique way of perceiving the world, although, there is still not enough designer doing Human-Centered AI research. Furthermore, that number is practically null when talking about particularly the ethics of AI, which is a field where designers could bring huge benefits.

chapter 10

Conclusion & Personal Reflection

This chapter explores a final conclusion of the project. This project was executed in the context of MOBGEN | Accenture Interactive and focused on the creation of support tools for a responsible development of AI application so they could also provide an added value proposition to their customers. Furthermore, as a final note, the project is finalized with a personal reflection regarding the personal development goals and the project in general.

10.1 CONCLUSION OF THE PROJECT

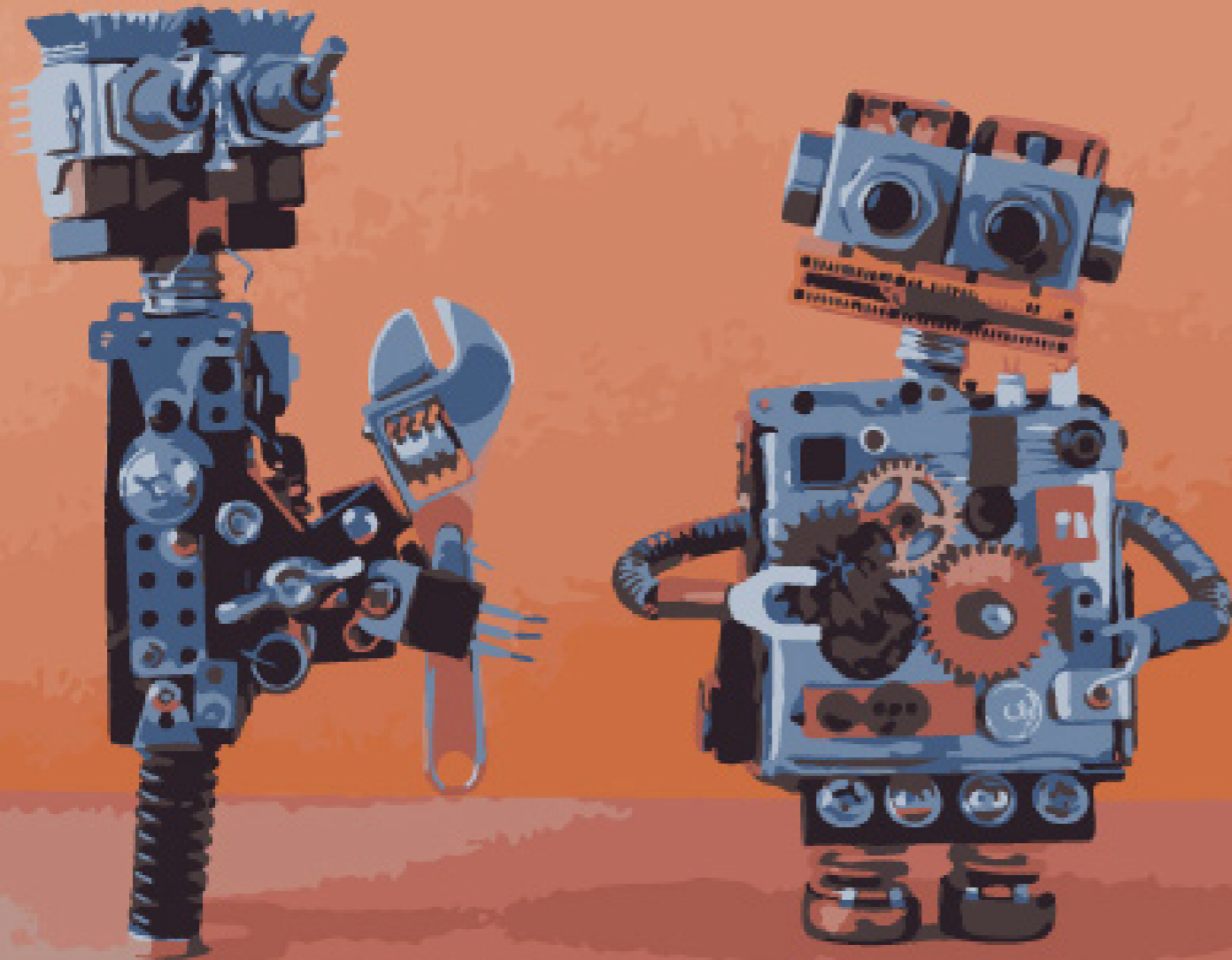
The purpose of this thesis was to design ethical support tools in the form of an ethical toolkit for MOBGEN | Accenture Interactive in order to apply them during the design process of an AI application. During this project it has been stated that an ethical assessment should be made on the development of AI applications. This has been concluded after a deep literature review stage where the fields of ethics and AI were investigated. Moreover, in order to understand more about the current context of the ethics of AI, a series of empirical studies were executed which included the organization of semi-structured interviews, as well as a creative discussion and the presentation of a “provotype” during an “Ethics & Design” seminar.

To achieve this, an ethical framework was created and translated into a canvas so that the organization (and other interested stakeholders) can use it to explain ethics and reflect about the ethics involved in the development of an AI application. Furthermore, this canvas features a puzzle that aims to explain in a playful manner the role and ethical impact that all the stakeholders involved in the process have. The insights gathered with the framework were translated to design principles that assisted with

the creation of an ethical strategy. Consequently, this strategy became an actionable ethical toolkit composed by 7 different modules distributed in 2 main ethical stages. The toolkit aims to trigger and support dialogue around AI ethics and to assist development teams in the ethical assessment of the development of AI applications.

Finally, the complete ethical toolkit was validated only once, however, its modules were tested individually in several occasions with both internal and external stakeholders. These validation sessions provided interesting feedback comments which helped in the redefinition and optimization of some of the modules. The results of this project are tangible and usable design tools that enable designers and developers of MOBGEN | Accenture Interactive, and other important stakeholders like AI researchers, to ethically assess the development of AI applications.

10



10.2 PERSONAL REFLECTION

At the beginning of this project I thought that I was familiar with the diverse concepts found in ethics and AI. However, I got an “uncomfortable” surprise to find out that I ignored more about ethics than I expected. Moreover, I shared the same level of understanding than most of the people that I got in contact with during the project. This made me think about the poor level of understanding we have about the topic and the risks that this represents today. We are approaching to an era where everything would be completely automated, hence, we have to be sure that AI would be a good reflection of us, an ethical machine.

At the beginning of the project I often got comments like “yeah, but you know, what is ethical and what is not is quite subjective”. This made me think about how people tends to reduce the level of importance of ethics since it is perceived as a limitation, mostly in the Engineering world. Furthermore, the lack of ethical understanding from most of the people I got in contact during the project motivated me to dive deeper into the concept and try to conceive a tool that could help designers bring this topic to the table. I have to say that I am satisfied with the result as I did not only designed an ethical tool to support critical reflection within Accenture and its dependencies, but also I created a little bit of awareness of this important topic by providing a value proposition out of the ethics of AI. This way, I hope that ethics could be considered more like an innovation engine more than a negative term.

During the course of this project, I experienced an academic and corporate environment in the quest of spreading the word about ethics, more specifically, the ethics of AI. In this trip I found myself connecting with a lot of highly intelligent people within both the corporate world, and the academic field. Furthermore, my networking efforts reached other Accenture dependencies like FJORD - The Dock located in Dublin, Ireland. Being an intern in MOBGEN | Accenture Interactive helped me also

to get in contact with a considerable big network of other interns of Accenture with similar topics (Human side of AI) that where more than excited to collaborate with me, and in the end we did by organizing the first “Human+AI College”.

Additionally, I am currently collaborating with an AI design team of FJORD - The Dock in order to expand the toolkit and implement it within their processes. I have to say that this happened because of Accenture’s global reach. I am sure that in a smaller company, this type of impact would not have been possible.

Regarding my personal development goals, I want to say that this project took me closer to my goal of becoming the bridge between AI developers, designers, and in this case ethicists as well. The IDE Faculty as helped me to build this skills and to put them on practice in a real-life setup, with this project being an example. The skills I practiced included preparing and performing interviews, facilitating creative discussions and workshop. Some of the most important things I learned about these skills is that there is no right or wrong way of doing them. The results depend completely on the direction that the interviewer or facilitator take on the execution of these.

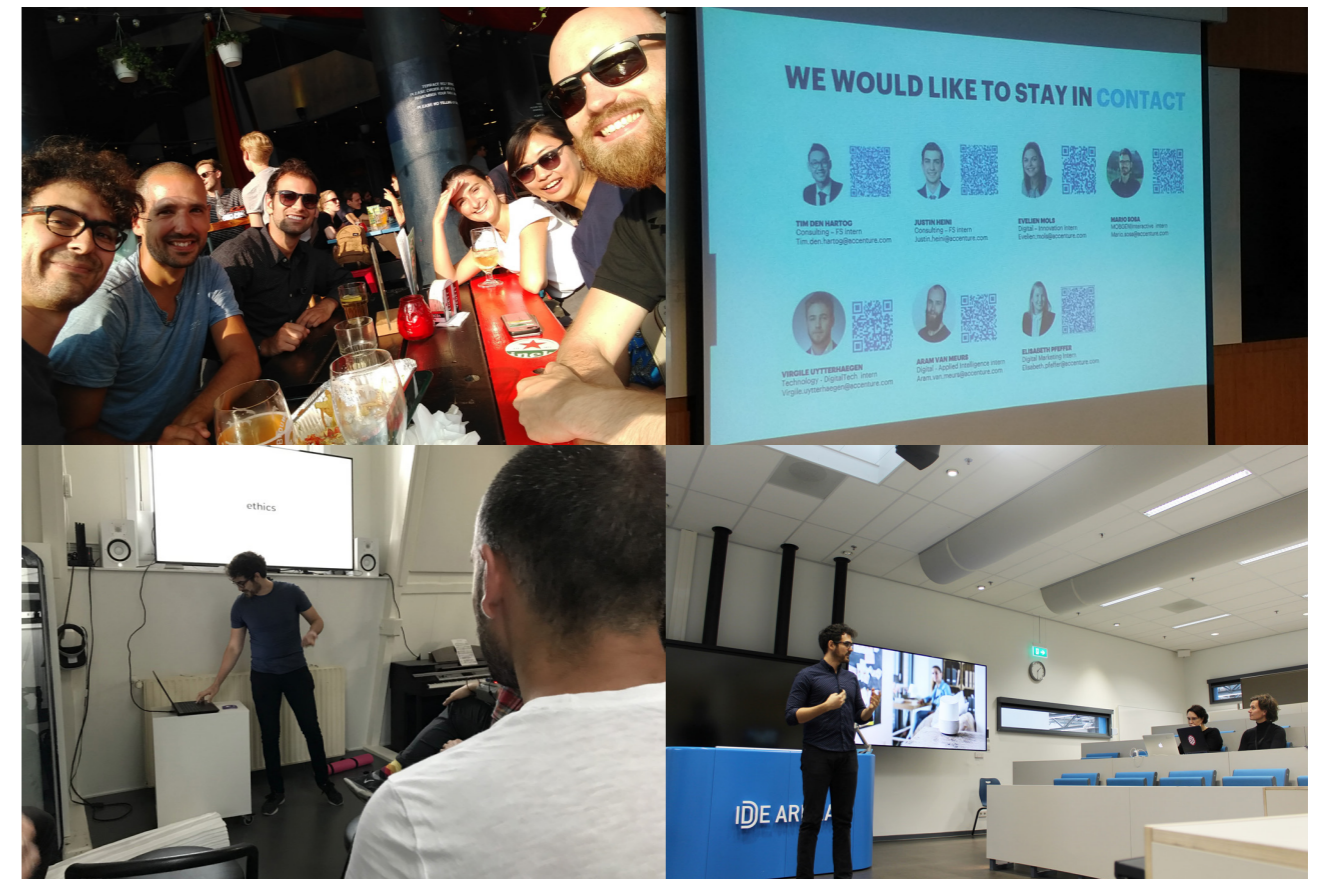
Certainly, I also practiced qualitative and quantitative analysis of data, which was something that I truly enjoyed in this project. Furthermore, I found it challenging to face the translation of this insights into a future vision and a tangible strategy that also could make an impact in an organization like MOBGEN | Accenture Interactive in an unexplored field as it is the Ethics of AI. Moreover, I gained knowledge on topics like ethics, the current applications and latest developments of AI, design led-innovation, digital transformation and the possible consequences of it. Thanks to the support of MOBGEN | Accenture Interactive I also improved my communication skills as I had to present my

progress continuously in front of a professional audience of designers, developers, and managers.

This project represents my personal contribution to the topic of design, ethics, and AI. I believe that design processes and tools can bring serious benefits to corporations like MOBGEN | Accenture Interactive, although, they might seem confusing and too complex at first. It can be observed in the validation stages of this project that these tools can generate discussion and switch people’s minds for good.

I am very curious about the future challenges what the future of design in ethics and AI will bring. But I am sure about one thing, I want to explore and face them in the future.

Finally, I want to say that I enjoy collaborating with MOBGEN | Accenture Interactive. I truly hope to have triggered in others the interest of exploring the field of the ethics of AI.



References

A

Accenture (2016). Accenture Acquires MOBGEN Expanding End-to-End Digital Services. Retrieved from <https://newsroom.accenture.com/news/accenture-acquires-mobgen-expanding-end-to-end-digital-services.htm> on 18-06-2019

Accenture (n.d.) Retrieved from <https://www.accenture.com/il-en/insight-artificial-intelligence-future-growth> on 18-06-2019

Accenture. (2018). Accenture Annual Report 2018 . Retrieved from <https://www.accenture.com/us-en/company-annual-report> on 18-06-2019.

Accenture. (2019a). Accenture Code of Business Ethics. Retrieved from <https://www.accenture.com/us-en/company-ethics-code> 18-06-2019.

Accenture (2019b). WEF 19 | Is Responsible AI Good Business? [Video file]. Retrieved from <https://www.youtube.com/watch?v=1SwwexNw9jo>

Amershi, S., Weld, D., Vorvoreanu, M., Fournay, A., Nushi, B., Collisson, P., ... & Teevan, J. (2019). Guidelines for Human-AI Interaction.

APM (n.d.). Ethics frameworks. Retrieved from <https://www.apm.org.uk/body-of-knowledge/people/professionalism/ethics-frameworks/> on 12-07-2019

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... & Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729), 59.

B

Barr, A. (2015) "Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms By Alistair" *The Wall Street Journal – Digits*. Barr, A. (2015) "Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms By Alistair" *The Wall Street Journal – Digits* . <https://blogs.wsj.com/digits/2015/07/01/google-mistakenly-tags-black-people-as-gorillas-showing-limits-of-algorithms/> .Retrieved on 18-06-2019.

Beeri, I., Dayan, R., Vigoda-Gadot, E., & Werner, S. B. (2013). Advancing ethics in public organizations: The impact of an ethics program on employees' perceptions and behaviors in a regional council. *Journal of business ethics*, 112(1), 59-78.

Bilderberg. (2019). 67th Bilderberg Meeting. Retrieved from <https://www.bilderbergmeetings.org/meetings/meeting-2019/press-release-2019> on 20-06-2019

Blasingame, J. (2017). When Trust Is A Best Practice, Profit Margins Increase. Retrieved from <https://www.forbes.com/sites/jimblasingame/2016/10/10/when-trust-is-a-best-practice-profit-margins-increase/#682df58e3bbe> on 10-07-2019

Bonde, S., & Firenze, P. (2013). Making choices: A framework for making ethical decisions.

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence*, 316, 334.

Burke, B. (2013). The Gamification of Business. Retrieved from <https://www.forbes.com/sites/gartnergroup/2013/01/21/the-gamification-of-business/#4a3a3144dc2c> on 09-07-2019

C

Chekanov, K., Mamoshina, P., Yampolskiy, R. V., Timofte, R., Scheibye-Knudsen, M., & Zhavoronkov, A. (2017). Evaluating race and sex diversity in the world's largest companies using deep neural networks. *arXiv preprint arXiv:1707.02353*.

Chowdhury, R. (2019). An AI governance approach to support innovation. Retrieved from Chowdhury, R. (2019). An AI governance approach to support innovation. Retrieved from <https://www.accenture.com/us-en/insights/artificial-intelligence/ai-governance-support-innovation>

Chui, M., Manyika, J., Miremadi, M., Henke et al. (2018). Notes from the AI frontier: Applications and value of deep learning. Retrieved from <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning> on 05-07-2019

Cutler, A., Pribic, M., & Humphrey, L. (2018). *Everyday Ethics for Artificial Intelligence*. IBM.

D

Davenport, T. H., & Ronanki, R. (2018). Artificial Intelligence for the Real World. *Harvard Business Review*, 96(1), 108-116.

Davenport,T., & Ronanki, R. (2018). Artificial Intelligence for the Real World. *Harvard Business Review*. Retrieved from <https://hbr.org/2018/01/artificial-intelligence-for-the-real-world> on 05-07-2019

Davis, D., O'Brien, M. A. T., Freemantle, N., Wolf, F. M., Mazmanian, P., & Taylor-Vaisey, A. (1999). Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes?. *Jama*, 282(9), 867-874.

DeMers, J. (2018). How Brands Should Be Working To Fix The Consumer Trust Crisis. Retrieved from <https://www.forbes.com/sites/jaysondemers/2018/05/09/how-brands-should-be-working-to-fix-the-consumer-trust-crisis/#3d98cf91554c> on 05-07-2019

Design Council. (2005). The 'double diamond' design process model. Design Council.

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue.

Dilam, I. (2016). My intership at Mobgen . Retrieved from <https://www.iremdilan.com/mobgen.html> on 26-07-2019

Dobson, J. (2003). Why ethics codes don't work. *Financial Analysts Journal*, 59(6), 29-34.

E

Ethics Guidelines for Trustworthy AI (European Commission, 2019); <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

Ethisphere (2019). Ethisphere Recognizes 128 World's Most Ethical Companies for 2019. Retrieved from <https://ethisphere.com/128-worlds-most-ethical-companies-for-2019/> on 08-07-2019

Étienne, B. (January, 2019) "Ten Issues to Watch in 2019," European Parliament- EPRS | European Parliamentary Research Service.

European Commission. (n.d.). Horizon 2020, Responsible research & innovation. Retrieved from <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation#Article> on 16-07-2019

Fjord (n.d.). Retrieved from <https://www.fjordnet.com/workdetail/cobe-the-chatbot-for-accenture-by-accenture/> on 25-06-2019

F

Fjord trends 2019. (2019). Retrieved from <https://trends.fjordnet.com/> on 18-06-2019

Fingas, J. (2016). Google AI finally loses to Go world champion . Retrieved from <https://www.engadget.com/2016/03/13/google-alphago-loses-to-human-in-one-match/> on 29-07-2019

Flasiński, M. (2016). *Introduction to artificial intelligence*. Springer.

Francis, D. (2017). Facebook's Ethical Problem. *Huffpost*. Retrieved from https://www.huffpost.com/entry/facebook-ethical-problem_b_59c923f9e4b0b7022a646c610n 04-07-2019

G

Gerdes, J. C., & Thornton, S. M. (2015). Implementable ethics for autonomous vehicles. In *Autonomes fahren* (pp. 87-102). Springer Vieweg, Berlin, Heidelberg.

Google. (2019). People + AI Guidebook. Retrieved on 20-05-2019 from <https://pair.withgoogle.com/>

Green, B. (n.d.). What is Technology Ethics? Markkuka Center for Applied Ethics. Retrieved from <https://www.scu.edu/ethics/focus-areas/technology-ethics/> on 10-07-2019

H

Halskov, K., & Dalsgård, P. (2006, June). Inspiration card workshops. In *Proceedings of the 6th conference on Designing Interactive systems* (pp. 2-11). ACM.

Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does Gamification Work?-A Literature Review of Empirical Studies on Gamification. In *HICSS (Vol. 14, No. 2014, pp. 3025-3034)*.

Hope, R. (2016). Why Microsoft's 'Tay' AI bot went wrong Retrieved from <https://www.techrepublic.com/article/why-microsofts-tay-ai-bot-went-wrong/> on 26-07-2019

I

Internet Encyclopedia of Philosophy, n.d. Ethics. Retrieved from <https://www.iep.utm.edu/ethics/> on 25-06-2018

J

Jill, T. (2017). World's biggest banks face £264bn bill for poor conduct. Retrieved from <https://www.theguardian.com/business/2017/aug/14/worlds-biggest-banks-face-264bn-bill-for-poor-conduct> on 10-07-2019

K

Kumar, J. (2013, July). Gamification at work: Designing engaging business software. In *International conference of design, user experience, and usability* (pp. 528-537). Springer, Berlin, Heidelberg.

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Sage.

L

Levin, S. (2018). Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian. Retrieved on 12-05-2018. <https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe>

Lockton, D. (2010). *Design with intent: 101 patterns for influencing behaviour through design*. Equifine.

Lubbock, H. & Virdee, J. (2017) "Live Innovation: Exploring the ethics of AI with WIRED Live" Retrieved from <https://www.fjordnet.com/conversations/live-innovation-exploring-the-ethics-of-ai-with-wired-live/> on 01-02-2019

Luppini, R. (Ed.). (2010). *Technoethics and the Evolving Knowledge Society: Ethical Issues in Technological Design, Research, Development, and Innovation: Ethical Issues in Technological Design, Research, Development, and Innovation*. IGI Global.

M

Makagon, K. (2019). The Growth And Potential Of AI. Retrieved from <https://www.forbes.com/sites/forbestechcouncil/2019/02/04/the-growth-and-potential-of-ai/#35e5a1fb7d7b>

Matney, L. (2018). Where Facebook AI research moves next. Retrieved from <https://techcrunch.com/2018/12/05/where-facebook-ai-research-moves-next/> on 18-06-2019

Mckinsey (2019). The ethics of artificial intelligence. Retrieved from <https://www.mckinsey.com/featured-insights/artificial-intelligence/the-ethics-of-artificial-intelligence> on 18-06-2019

McMillan, M. (2012). Codes of Ethics: If You Adopt One, Will They Behave? Retrieved from <https://blogs.cfainstitute.org/investor/2012/02/20/codes-of-ethics-if-you-adopt-one-will-they-behave/> on 16-07-2019

Mitchell, J. (2018). How AI, Machine Learning And Other Disruptive Trends Are Defining The Future Of Customer Service. Retrieved from <https://www.forbes.com/sites/julianmitchell/2018/09/05/how-ai-machine-learning-and-other-disruptive-trends-are-defining-the-future-of-customer-service/>

trends-are-defining-the-future-of-customer-service/#38e829ff4cdf on 18-06-2019

Mobgen. (2019). Artificial Mirror, Training a neural network with your emotions. Retrieved on 20-05-2019 from <https://www.mobgen.com/portfolio-item/artificial-mirror/>

Mobgen | Accenture Interactive (2019). Accenture Interactive Amsterdam opens the doors to Building 01, positioning itself as The Experience Agency in The Netherlands. Retrieved from <https://www.mobgen.com/accenture-interactive-amsterdam/> on 26-07-2019

Moriarty, J. (2017). "Business Ethics", *The Stanford Encyclopedia of Philosophy (Fall 2017 Edition)*, Edward N. Zalta (ed.),

N

Noessel, C. (2017). *Designing agentic technology: AI that works for people*. Brooklyn, NY: Rosenfeld Media.

Norberg, P. (2009). "I don't care that people don't like what I do"—Business codes viewed as invisible or visible restrictions. *Journal of Business Ethics*, 86(2), 211-225.

P

Paul, K. (2019). 'Disastrous' lack of diversity in AI industry perpetuates bias, study finds. *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2019/apr/16/artificial-intelligence-lack-diversity-new-york-university-study> on 01-07-2019

Personal Data Protection Commission Singapore (PDPC) (2018). DISCUSSION PAPER ON ARTIFICIAL INTELLIGENCE (AI) AND PERSONAL DATA – FOSTERING RESPONSIBLE DEVELOPMENT AND ADOPTION OF AI. Retrieved from <https://www.pdpc.gov.sg/Resources/Discussion-Paper-on-AI-and-Personal-Data> on 09-07-2019

Peters, A. (2018). This tool lets you see—and correct—the bias in an algorithm. Retrieved from <https://www.fastcompany.com/40583554/this-tool-lets-you-see-and-correct-the-bias-in-an-algorithm> on 18-06-2019

Polachowska, K. (2019). Tinder algorithms: how the matchmaking happens. Retrieved from <https://neoteric.eu/blog/tinder-algorithms-how-the-matchmaking-happens/> on 05-07-2019

Popper, K. (2012). *The open society and its enemies*. Routledge.

Price, R., Wrigley, C., & Matthews, J. (2018). Action researcher to design innovation catalyst: Building design capability from within. *Action Research*, 1476750318781221

Purdy, M., & Daugherty, P. (2018). How AI Boosts Industry Profits & Innovation. Accenture. Retrieved from https://www.accenture.com/_acnmedia/PDF-84/Accenture-AI-Industry-Growth-Full-Report.pdf on 05-07-2019

R

Reason, P., & Bradbury, H. (Eds.). (2001). *Handbook of action research: Participative inquiry and practice*. London: Sage Publications.

Reisinger, D. (2019). Expert says that AI can replace 40% of jobs in 15 years. Retrieved from <https://fortune.com/2019/01/10/automation-replace-jobs/> on 09-07-2019

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016, March). Overtrust of robots in emergency evacuation scenarios. In *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on* (pp. 101-108). IEEE.

S

Saffer, D. (2005). *The role of metaphor in interaction design*. Information Architecture Summit, 6.

Sanders, E. B. N., & Stappers, P. J. (2014). Probes, toolkits and prototypes: three approaches to making in codesigning. *CoDesign*, 10(1), 5-14.

Schwab, K. (2014). The profitability of trust. Retrieved from <https://www.weforum.org/agenda/2014/12/the-profitability-of-trust/> on 10-07-2019

Senges, M., Ryan, P. S., & Whitt, R. S. (2017). *Composite Ethical Frameworks for IoT and other Emerging Technologies*. Available at SSRN 3092362.

Senges, M., Ryan, P. S., & Whitt, R. S. (2017). *Composite Ethical Frameworks for IoT and other Emerging Technologies*. Available at SSRN 3092362.

Singh, J. J., Iglesias, O., & Batista-Foguet, J. M. (2012). Does having an ethical brand matter? The influence of consumer perceived ethicality on trust, affect and loyalty. *Journal of business ethics*, 111(4), 541-549.

Socher, R. (2019). *Why Ethical AI Is A Critical Differentiator*. Retrieved from <https://www.forbes.com/sites/insights-intelai/2019/03/27/why-ethical-ai-is-a-critical-differentiator/#9fd012c63ab7> on 18-06-2019

T

Tassoul, M. (2012). *Creative facilitation*. VSSD.

Tyagarajan, N. (January, 2019). *Changing the AI narrative – a call to action*. Retrieved from <https://www.weforum.org/agenda/2019/01/changing-the-ai-narrative-a-call-to-action/> on 20-06-2019

U

Universidad de Salamanca (2018). *Directrices éticas para una Inteligencia Artificial fiable*. Retrieved from <https://universoabierto.org/2018/12/30/directrices-eticas-para-una-inteligencia-artificial-fiable/> on 28-07-2019

V

Van de Poel, I., & Royakkers, L. (2011). *Ethics, technology, and engineering: An introduction*. John Wiley & Sons.

Van der Beek, P. (2019). *AP pleit voor openheid bij overheid over algoritmes*. Retrieved from <https://www.computable.nl/artikel/nieuws/big-data/6673350/250449/ap-pleit-voor-openheid-bij-overheid-over-algoritmes.html> on 20-06-2019

Veruggio, G. (2006, December). *The euron roboethics roadmap*. In *Humanoid Robots, 2006 6th IEEE-RAS International Conference* on (pp. 612-617). IEEE.

Victor, D. (2016). *Microsoft Created a Twitter Bot to Learn From Users. It Quickly Became a Racist Jerk*. Retrieved on 25-05-2018. <https://www.nytimes.com/2016/03/25/technology/microsoft-created-a-twitter-bot-to-learn-from-users-it-quickly-became-a-racist-jerk.html>

Vincent, J. (2019). *THE PROBLEM WITH AI ETHICS*. Retrieved from <https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech> on 08-07-2019

Visser, F. S., Stappers, P. J., Van der Lugt, R., & Sanders, E. B. (2005). *Contextmapping: experiences from practice*. *CoDesign*, 1(2), 119-149.

W

Webley, S., & Werner, A. (2008). *Corporate codes of ethics: Necessary but not sufficient*. *Business Ethics: A European Review*, 17(4), 405-415.

Z

Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. Available at SSRN 3312874.

Zhou, J. (2018). *The Machine Learning Ethical Toolkit*. Retrieved from <http://www.jamesxzhou.com/work/machine-ethics-toolkit/> on 10-10-2019

