



CERTIFIED

APPROACH TO HUMAN-CENTERED AI ADOPTION

A framework for agile policymaking to regulate technology

Master Thesis
Michelle Geerlings

Master thesis - Strategic Product Design

Faculty of Industrial Design Engineering
Delft University of Technology

The CERTIFIED approach to human-centered AI adoption
A framework for agile policymaking to regulate technology*

* Creative, Empathic, Reflective, Transparent, Inclusive, Future-oriented, Iterative, Experience-driven, Demonstrative

Defended on 18 December 2020

Written and designed by
TU Delft supervision by

Michelle Geerlings
Deborah Nas
Jeroen van Erp

Dutch Ministry of Justice and Security
supervision by

Joël van Andel



Ministerie van Justitie en Veiligheid



PREFACE

In front of you lies the end result of my graduation project for the master Strategic Product Design at Delft University of Technology. This thesis describes the activities performed at the Ministry of Justice and Security in the period April-December 2020. The thesis describes the challenge of the ministry to regulate the rapid pace of digital technology development. The scope of this thesis is the study of an agile policymaking process for Artificial Intelligence (AI) technology. An explorative study is conducted in the intersection of the design, AI and political context. This intersection resembled how design practices could enhance organizational learning in its approach to solve complexity and uncertainty raised by AI applications. Several AI cases have been studied that supported the research in making the dialogue more explicit. Due to the Corona pandemic, most of the work has been done through online meetings.

Design Thinking as a scientific, human-centered, and empathetic process that gets us to a platform of understanding the people and their problems with firm understanding and empathy, along with continuous innovation. And as a designer, I therefore strongly believe a human-centered approach is essential in each design project. This process helps uncover the needs of the target group which can be translated into useful design elements. This phenomena is in particular challenging in the political context. At any moment in any social system there are multiple, competing and even incompatible interests and thus visions of what is the preferred situation. This makes it challenging to grasp latent needs of the endless variety of people represented in the social system, of which the government in the end is held accountable. In order to design a meaningful solution for the societal complex challenges policymakers face, I took an active part in the team to experience their challenges myself. With this thesis, I hope to inspire policymakers with my human-centered design approach.

Enjoy reading!

Michelle

ACKNOWLEDGEMENTS

In this section I would like to express my gratitude to several people that helped me during this graduation project. First of all, I want to thank my TU Delft supervisors. Jeroen, thank you for challenging me in this project and for always pointing me on interesting and promising options to include in my research and design. Deborah, I want to thank you for your endless support during the project, for empowering me when I presented my ideas and for being critical to give them more shape. But most of all, I appreciate you both helped me to ground, and for letting me feel comfortable exploring the unknown. During our two-weekly 'dreamteam' sessions I learned to make sure all elements of the vision and toolkit connected and that they should suit the narrative. I want to thank you both for being there for me during this project, both physical and due to the corona pandemic mainly online. Although this was the case, I knew I could always count on your involvement and the support in the times I felt lost.

Furthermore, I would like to thank the AI team to give me the opportunity to conduct this explorative study within such a warm and inspiring team. Special thanks to Joël, my supervisor at the Ministry of Justice and Security, for the support and valuable insights you provided me with. I enjoyed showing you different design approaches and it gave me energy to see that you got inspired by these new methods. Thanks to Michel, my director, for your energy, your open mindset, and the opportunity to express my ideas on using different strategies and design methods that shaped the base of the AI team today. Olof, thank you for the inspiring conversations during our coffee breaks, your critical views and your positive words throughout the project. Caspar, Michael, Eline, Jorien, Justyna, and Natalie, I really enjoyed our weekly meetings, virtual coffee breakouts and strategy sessions. They provided me with so much energy to embrace this challenge with both hands. I'm grateful to see that the proposed design methods and programmes have become a solid foundation for the way of working. Let our team be the inspiring pioneers that other teams within the ministry would like to become themselves.

I would like to thank my family and friends for their endless support throughout the process. In particular Regina, Pamela, Eva, Tirza, Mariska and Marie for the many joyful moments in my graduation journey during this COVID-19 period.

Lastly, I would like to thank everyone that took their time to participate in all the different activities from online validations to physical sparring moments. Thank you for sharing your knowledge in the areas of ethics, journalism, philosophy, IT, legislation, systemic design, data science, organizational structures, and challenges in startup collaborations. These valuable sparring moments gave me insights in the complex social system and the interests and viewpoints to take into account. It means a lot to me that you felt comfortable to share your personal experiences or provide me with honest feedback.

EXECUTIVE SUMMARY

The effects of Artificial Intelligence have been widely expressed in the news. Augmenting our intelligence with artificial intelligence has the potential of helping civilization flourish like never before. However, AI also has a downside. The discussion about privacy versus security has been raging for some time. Deep fakes make us question what information we see is real. (Unintended) biased systems evoke discrimination as they violate our human rights. Even criminality has taken a more digital role (e.g. AI-driven hacking, identity theft, fraud), challenging the government in its attempt to keep up with the digitalization pace regarding justice and public safety. Now that technology can no longer be ignored in our lives and our current approach to safeguard public safety won't be enough, the question is: how do we manage to keep the technology beneficial?

The Dutch Ministry of Justice and Security stated that policy development is threatened to lag behind regarding the regulation of Artificial Intelligence applications, and stated its desire to catch up. Yet, 95% of current policymaking is swayed by the issues of today (e.g. a slip of the tongue of a minister). Both empirical and literature research revealed that *"a vision of facial recognition will not come until the newspaper has published that the police are experimenting with it."* This represents a cycle of reactive policymaking, as it is driven by external pressure.

As the media serves as the magnifying glass on the government, the pressure on the policymakers to intervene increases. The longer we wait, the wider the 'gap' will become between regulation and market development. The Big Tech already start to assume the role of regulator introducing sets of principles and ethical policies. If the government does not catch up with interventions, the power concentration will shift to those who are AI-savvy. The importance of autonomy is under pressure and the question if it's legally allowed is not the end of the discussion. It is time for proactive policymaking, by openly questioning ourselves: What do we as society really desire? And is AI perceived as proportionate to the stated goal?

The Overton Window reveals that the political viability of policy interventions mainly depends on whether it falls within the range of acceptability. The COVID-19 pandemic showed us a window of opportunity. In times of crisis, we tend to have a wider range of policy acceptability. As the government responds to the Covid-19, it's preferred styles of action are changing. Preferable futures people may consider today, could be more divergent and more radical than before. The key insight is that people want to know what they are up to.

Hence, it is time to make explicit which values outweigh others and retain control of our future relationship with AI technology. Based on a clear vision of the future, contemporary choices can be weighed up (i.e. on whether and how facial recognition technology should be employed by domestic law enforcement agencies).

In order to keep this technology beneficial, a human-centered approach should be key not only in AI development, but also in the process of policymaking. Instead of focussing on 'what is possible', we need to empathize with identified stakeholders to learn 'how to unravel the latent needs defining society's desires of today'. Along with the unfamiliarity of AI technology and its unpredictable negative consequences, it requires a transition towards becoming an adaptive and resilient government. This transition represents a shift in culture (from 'assuming' to 'learn by doing'), in mindset (from 'is it safe?' to 'is it safe to fail?'), and in the current organizational approach (from reactive to proactive policymaking).

Proactive policymaking is recognized as a **future-oriented** exploration. To unravel the desirability of AI applications in public services, *AI experiments should be framed as a priority vehicle for organizational learning*. Experiments are used to **empathize** and gain feedback early in the process. *Policy Prototypes* and *Minimum Viable Policies* are created as first feedback loops prior to current feasibility studies. This involves a **creative** and **iterative** way of working, by proposing policy options to a representative group of society early in its development. An 'Desirability panel', covering the **experiences** of the Quadruple Helix (government, business, academia and society), reduces the blind spots and offers **transparency**. It facilitates the short-line communication and alignment from the relevant stakeholders. Potential policy interventions transcend theoretical discussion as the **demonstrated** experiments provide a context for detailed dialogue. Tangible examples help to demystify the fuzziness of what society defines as 'ethically responsible'. The daily activities of a policymaker will not be swayed by the issues of the day anymore as stakeholder **inclusion** is at the heart of policy development. Hence, the policymaker is less affected by the magnifying glass of the media.

This vision calls for the need of a **Creative, Empathic, Transparent, Inclusive, Future-oriented, Iterative, Experience-driven** and **Demonstrative** approach towards policymaking to steer human-centered AI adoption. Therefore I present the acronym: **CERTIFIED**. The CERTIFIED approach covers a theoretical framework for agile policymaking. In essence, it is about the ability to adapt quickly by providing a window for feedback by the public early-on in the process and embracing criticism as a valuable gift.

An iterative design process led to (1) the creation of the CERTIFIED manifesto for the ministry, that publicly declares the needed mindset in a set of principles, and (2) a practical toolkit to offer policymakers an actionable way to carry out these principles.

The manifesto represents an empowering guide for the transition. In order to activate the CERTIFIED manifesto, a toolkit resembles a low barrier to experience taking the first step in demystifying the desirability of AI applications with a representative group of society. This toolkit contains a Perspectives card deck and a set of canvases. The Perspectives carddeck include 15 stakeholder cards, representing their values, roles, concerns and power. This card deck offers a structural approach for policymakers to emphasize with the stakeholders affected by the proposed AI systems. The set of canvases follow the structure Context - Empathy - Define - Ideate- Minimum Viable Policy - Test, in separate canvases to guide the panel in using Design Thinking. These 6 canvases structure the use of this toolkit and capture the insights from the dialogue.

A final pilot with representative stakeholders of the Quadruple helix showed a cross-fertilization of knowledge and expertise resulting in the ability to identify their blind spots and to formulate a 'Minimum Viable Policy' and test set-up to validate the assumptions on its perceived desirability. This research concludes that if space is provided for proactive policy experimentation, this would give policymakers momentum to reflect on the long-term goals instead of the issues of the day, thus offering a way towards becoming a resilient government with agile policymaking.

You have taken the first step by reading this thesis to trigger your curiosity and spark your thinking. I hope to inspire you to become the pioneer that will guide others towards an adaptive and resilient government. It's time to change, this is your call.

READING GUIDE

The reading guide exhibits the overview of the report to assist the reader towards the desired information. At the right page is a short description per chapter which discloses its content.

Each chapter starts with a short introduction after which the topic and conducted research is elaborated upon. At the end of the main section is an overview of the key insights on a purple background. These represent the building blocks and rationale for the final design in the recommendations for the Dutch Ministry of Justice and Security.

Text

Bold text

Bold text indicates the key message

Italic text

Italic text refers to quotations or definitions

Conclusions

Text and visuals on the purple background contain the most crucial insights of the design and recommendations

Examples

Illustrative examples and quotes for a deeper understanding are written or visualized on the green background

DEFINITIONS

This thesis covers both literature and internal research, in which technical, political and design jargon are used. To avoid complexity and misunderstanding the definition of some jargon terms are clarified in relation to the research scope.

Policy *The chosen format of (governmental) intervention to achieve a stated goal.*

Policymaking *The process of defining the right suited code of conduct or action to achieve a (mostly political) stated goal.*

Agile *(1) ability to move quickly and easily; (2) a project management method that is characterized by the division of tasks into short phases of work and frequent reassessment and adaptation of plans.*

Minimum Viable Product *A product with (just) enough features to validate a product idea early in the product development cycle. The MVP can help the creator receive user feedback as quickly as possible to iterate and improve the product.*

Minimum Viable Policy *A policy draft with (just) enough features to validate the acceptability rate and effectiveness early in the process of policymaking.*

Provotypes *A digital or physical design artefact whose main goal is to provoke discussion among different types of users and stakeholders. The artefact is one step prior to the Prototype, but transcends a theoretical discussion.*

Prototypes *An early sample or model built to test a concept or process. This sample is generally used to evaluate a new design to identify potential errors at an early stage.*

Pilot *A small scale preliminary study conducted in order to evaluate feasibility, duration, cost, adverse events, and improve upon the study design prior to performance of a full-scale research project.*

Sprint *A short, time-boxed period set for a team to complete an aligned amount of work.*

Design Thinking *A non-linear, iterative process that is applied to understand users, challenge assumptions, redefine problems and create innovative solutions to prototype and test. Involving five phases—Empathize, Define, Ideate, Prototype and Test—it is most useful to tackle problems that are ill-defined or unknown*

Manifesto *A short published declaration that “manifests” or promotes a set of intentions, motives and goals. A manifesto is mainly considered passionate, compelling and vivid to gain support and incite to act.*

Toolkit *A set of tools designed to be used together or for a particular purpose*

Inclusion *This term is mostly associated with minority groups, but in this thesis I define inclusion as the act of including external perspectives into policy development. The minority here refers to other professional roles rather than solely traditional policymakers (i.e. Big Tech, startups, IT developers, journalist, philosophers etc.)*

TABLE OF CONTENTS

< PREFACE	4
< ACKNOWLEDGEMENTS	5
< EXECUTIVE SUMMARY	6
< READING GUIDE	8
< DEFINITIONS	9
01 INTRODUCTION	12
1.1 Project Context	13
1.1.1. The rise of AI in digital services	13
1.1.2. The need to act now	15
1.1.3. The role of the government	15
1.1.4. The AI team	16
1.2 Project Approach	18
1.2.1 Project aim	18
1.2.2 Research questions	18
1.2.3 Process approach	19
1.2.4 Literature scope	21
02 EXPLORATIVE STUDY	22
> VISUAL READING GUIDE	23
2.1 Deepdive in AI	24
2.1.1 Definition of AI	24
2.1.2 The components of intelligence	24
2.1.3 Value of AI applications	26
2.1.4 Main challenges of AI	27
2.1.5 The call for transparency	31
2.1.6 The rise of ethical tools	32
2.1.7. Insights & Conclusion	34
2.2 The political context	35
2.2.1 Tackling a wicked problem	35
2.2.2 The role of the media	38
2.2.3 Policymaking process	42
2.2.4 Agile policymaking	46
2.2.5. Insights & Conclusion	51
2.3 Current state of AI experimentation	52
2.3.1 Spectrum of experimentation	52
2.3.2 The lenses of governmental innovation	56
2.3.3 Identified challenges in AI experiments	58
2.3.4. Insights & Conclusion	61

03 PROBLEM REFRAMING	62
3.1 Revealing the key problem	63
3.1.1. Clustered problem landscape	63
3.1.2 Revised problem statement	64
3.2 A framework for agile policymaking	65
04 DESIGN VISION	67
4.1 Design vision (the why)	68
4.2 CERTIFIED manifesto (the how)	70
4.3 Toolkit Design (the what)	72
4.3.1. Reasoning for the creation of a toolkit	72
4.3.2 Toolkit design requirements	72
05 CONCEPT CREATION	73
5.1 Conceptualisation	74
5.1.1. Ideation	74
5.1.2. Iterative process	74
5.2 Final concept	75
5.2.1. The CERTIFIED toolkit	78
5.2.2. The setting	82
5.2.3. The implementation	84
06 CONCEPT VALIDATION	87
6.1 Evaluation pilot	88
6.2 Key insights	90
07 RECOMMENDATIONS & DISCUSSION	93
7.1 Discussion & Limitations	94
7.2 Recommendations	96
08 PERSONAL REFLECTION	98
8.1 Personal ambitions	99
8.2 Thesis result	100
REFERENCES	101
APPENDICES	105



01 INTRODUCTION

This chapter introduces the project context and describes the approach for the conducted research.

1.1 Project Context

This section describes the project context, the urgency to act now as government and the role of the AI team.

1.1.1. The rise of AI in digital services

In 2020, the world changed fundamentally, and so did our interdependence with data. As COVID-19 swept the globe, nearly every aspect of life moved online, and people depended more and more on apps and the internet to socialize, educate and entertain ourselves. We are living in an era of data explosion. *“Every day we create 2.5 quintillion bytes of data at our current pace, but that pace is only accelerating with the growth of the Internet of Things”* (Cath, 2018). Over the last two years alone 90 percent of the entire digital data in the world was generated. (ibid)

Artificial Intelligence (AI) is one of the game changers in this world of digital transformation. It is expected that In 2030, algorithms will play an important role in more and more everyday parts of our lives. Algorithms already influence how we spend our free time, how we find our way and how we choose our partner. Until now, however, the great advances in artificial intelligence have been mostly theoretical. The impact of these technological developments is expected to accelerate in the coming decades.

We see a trend where every appliance becomes ‘smart’, as it is embedded with intelligence. From smart streetlights to everyday appliances that respond to voice commands and anticipate to your needs. Appliances could order groceries, finetune light intensities, and could offer live stream recordings. Combined with the explosion of low-cost microscopic sensors and the deployment of high-bandwidth networks, we’re heading towards a trend where every device has the potential to become a ‘smart device’.

Where the AI maturity still is in the narrow AI stage, futurists believe AI will achieve human-level intelligence (also known as general AI or AGI). Therefore, the question is not if, but when AI will achieve this stage. According to a survey by AI Impacts shows the median prediction in statements is 2042, though predictions of AGI researchers and futurists have medians in the early 2030s. *“This first rough artificial general intelligence may well be created by around 2030 (25% chance), 2040 (50% chance) or never (10% chance).”* (McKinsey, 2020) However, experts note that it won’t be until 2060 until AGI has gotten good enough to pass a “consciousness test”. In other words, we’re probably looking at 40 years from now before we see an AI that could pass for a human.

A trend is recognized where machine learning scripts are increasingly made open source, available on the cloud, allowing any individual with an internet connection to supplement their cognitive ability, augment their problem-solving capacity, and build new ventures at a fraction of the current cost. Where this is now only reserved for the ones capable to code, this capability is growing among the young. Driven by the convergence of global high-bandwidth connectivity, neural networks, and cloud computing, every industry, from healthcare, education, to entertainment, will be impacted. The rise of “AI as a Service” (AIaaS) platforms will enable humans to partner with AI in every aspect of their work, at every level. AIs will become entrenched in everyday business operations, serving as cognitive

collaborators to employees—supporting creative tasks, generating new ideas, and tackling previously unattainable innovations. In some fields, partnership with AI might even become a requirement, like making certain diagnoses without the consultation of AI may be deemed malpractice.

News publications revealed how Clearview AI collected online photos, storing them in databases without consent of the people depicted in the images. (Hill, 2020) The internet, and therefore the data that fuels an AI, does not know regional borders. We've seen AI software influencing the elections targeting tailored information based on your search and interests to nudge you voting for a specific party. But also the rise of deep fakes, where the shift is made towards questioning what information we see is real and what isn't. Elon Musk is calling out AI as 'one of the biggest existential threats to human existence of all time' and states we need to regulate AI before it becomes a danger to humanity. Recommendation systems cater for filter bubbles. The selective provision of information increases the probability of giving rise to polarisation, as people strongly believe what information is presented to them online, entering a rabbit hole of disinformation. Predictive policing software that creates risk profiles of people. This aims to violate the human right "innocent until proven otherwise". What value should weigh higher than the other is a political debate. There is no right or wrong, but the choice to agree or disagree.

There is an evident need for such ethical guidelines since legal policy often lags behind technological progress, leaving an expansive grey area where technology creates opportunities for new practices that have yet to be challenged by law and regulation (Aizenberg & Van den Hoven, 2020). Moreover, the law does not regulate everything, leaving space for human actions to be guided by values rather than explicit laws. Not every unethical use of data is illegal, but can nevertheless be defined as undesirable in view of certain values or context (Van Schie, Westra, & Schäfer 2017). To give an example, while it is certainly legal to use certain types of surveillance measurements to gain insight into what employees do, it might not be ethical to do so. What thoughts elicit when you imagine a risk profile is made of you using facial recognition? Take a look at Figure 1: in what situation is this acceptable for you?

FACIAL RECOGNITION SOFTWARE

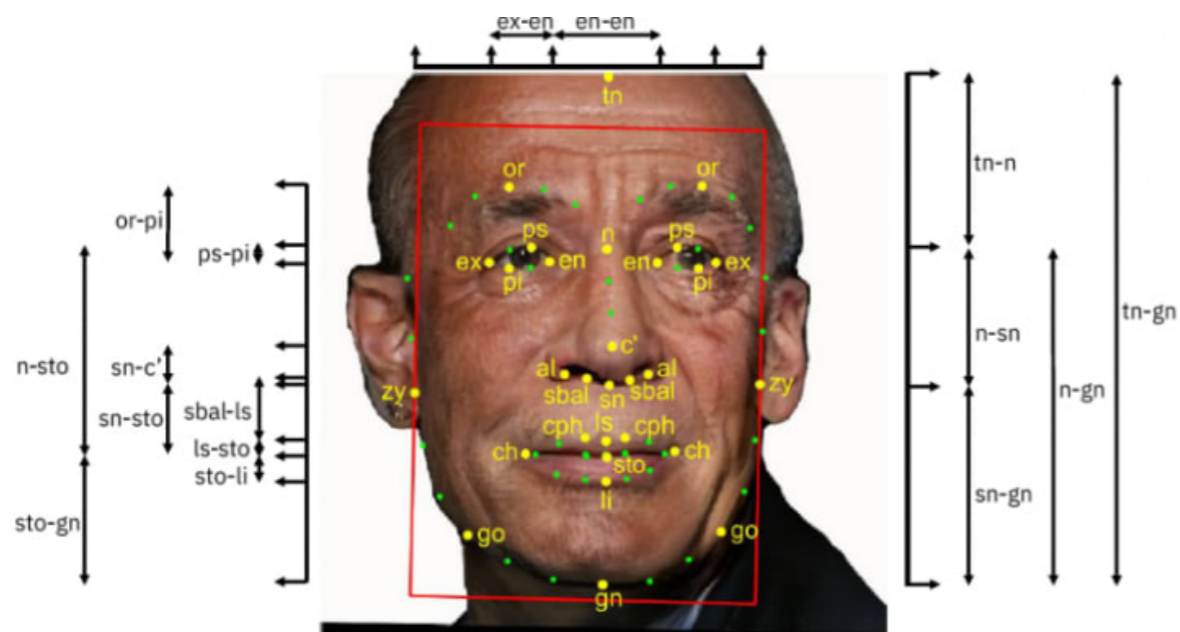


Figure 1: Illustrative representation of Facial Recognition software

1.1.2. The need to act now

“We believe now is the time to begin a national dialogue on whether and how facial recognition technology should be employed by domestic law enforcement agencies.” - Krishna (CEO of IBM, 2020)

AI offers great opportunities to improve our lives but such a disruptive technology also requires ethical and security-by-design considerations. The main challenge is that current legislation can't keep up with technological advancements by the market. The rise of unethical AI applications draws urgency to the need for laws and regulation to protect the human rights of citizens (i.e. new articles of ethnic profiling at tax authorities, court decision to quit SyRi (System Risk Indication software), the recently published EU whitepaper on ethical AI). In addition, criminality has taken a more digital role (i.e. cybercrime, identity theft, fraud) and the executing organizations (e.g. the court, the police) can't keep up with the digitalization pace to safeguard public safety. The government collects a vast amount of digital data on a daily basis and AI applications show potential to overcome the bureaucratic ties public organizations are caught up in. AI can assist police work in pro-active criminal activity prediction, real-time criminal activity detection, and time consuming manual activities such as filtering relevant information out of large datasets. However, AI can also be used for criminal purposes to detect weak spots in security systems and the unfamiliarity of the risks and impact on people's lives raises AI anxiety among citizens. Considering the use of algorithms on data brings about its legal, ethical and regulatory challenges. The use of data must take into account concerns regarding data security, potential issues caused by bias and protect human rights (i.e. privacy). Most people still have little idea of what is happening with their data. However, when people become aware that they are being watched, it can lead to self-censoring and risk-avoiding behavior. Thus, a broader engagement of civil society is needed on the values that need to be embedded in systems that use AI and the directions for future development.

The longer we wait, the wider the 'gap' will become between regulation and market development. Now is the time to make explicit which values outweigh others and retain control of our future relationship with this technology. We need to further define its added value and power to understand the consequences we must consider. Based on a clear vision of the future, contemporary choices can be weighed up and decisions can be made. Such as: Should we ban deepfakes? Big Tech already starts to assume the role of regulator introducing sets of principles and ethical policies. If the government does not catch up with interventions, the power concentration will shift to those who are AI-savvy. The importance of autonomy is under pressure. *“As artificial intelligence (AI) systems make more decisions, AI ethics become increasingly relevant to public policy.” (Deloitte, 2019)*

1.1.3. The role of the government

The government consists of the central government, the provinces and water authorities, and the municipalities, covering the national, regional and local focus. These authorities take measures, enact laws and monitor compliance. The central government consists of our 12 ministries in The Hague, the executive organisations that fall under these ministries and the High Councils of State.

One of these 12 ministries is the Ministry of Justice and Security, which is responsible for Justice, Imprisonment, and Public security. Hereafter 'ministry' refers to the Ministry of Justice and Security until stated otherwise. This ministry ensures the rule of law in the Netherlands,

so that people can live together in freedom, regardless of their lifestyle or views. They strive for a safer and fairer society by giving people legal protection and intervening in their lives where necessary. This ranges from drastic measures to opening up to new perspectives. It requires choices, of which the government weighs up interests, invests in the future and intervenes when necessary. (Rijksoverheid, 2020).

The ministry's board of Director-Generals noted that they threaten to lag behind policy developments in the field of AI and desire to catch up. Therefore, the board decided to establish an AI policy team. As of February 2020 the AI team is founded with their mission to supervise AI development and create a regulatory framework for human-centered AI adoption. (Kwartiermakersplan, 2019)

1.1.4. The AI team

The AI team started off with four members in April, and is now represented by 10 policymakers. The perceived ambition of the ministry in relation to AI is visualized in Figure 2.

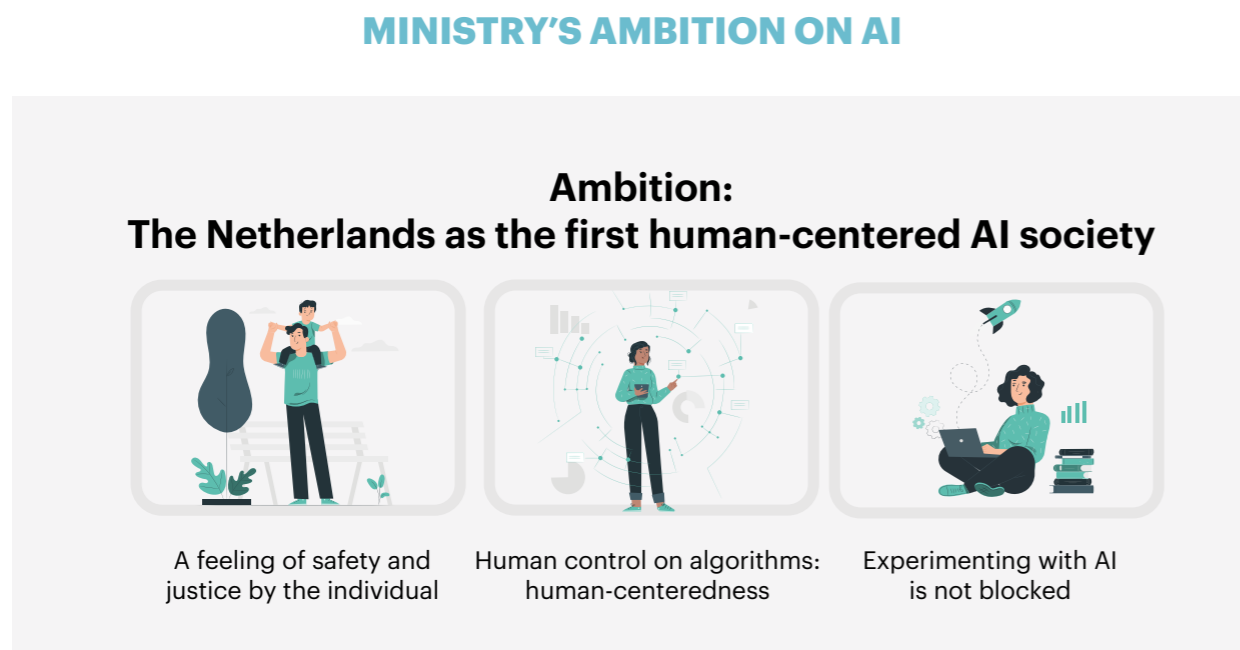


Figure 2: the ambition on AI of the ministry

The ambition is to make The Netherlands the first human-centered AI society. The emphasis on being the first prevails a push to fundamentally integrate AI into public services. Hence, quickly learn to use AI optimally by experimenting, so that the Dutch government is the first to be skilled in its practice. To realize human-centered AI, the government must protect people from the negative effects of AI, both preventively and reactively. This means being able to respond quickly to threats that AI can cause and identify unethical use and prevent damage at an early stage. This requires keeping the human in control. Therefore, the mission of the AI team is threefold: (1) optimize the chances AI has to offer, (2) mitigate the threats caused by AI applications, and (3) prevent unethical use of AI applications. A communication visual is created to express the role and activities of the AI team in relation to the ambition (see Figure 3).

ROLE AND ACTIVITIES OF THE AI TEAM

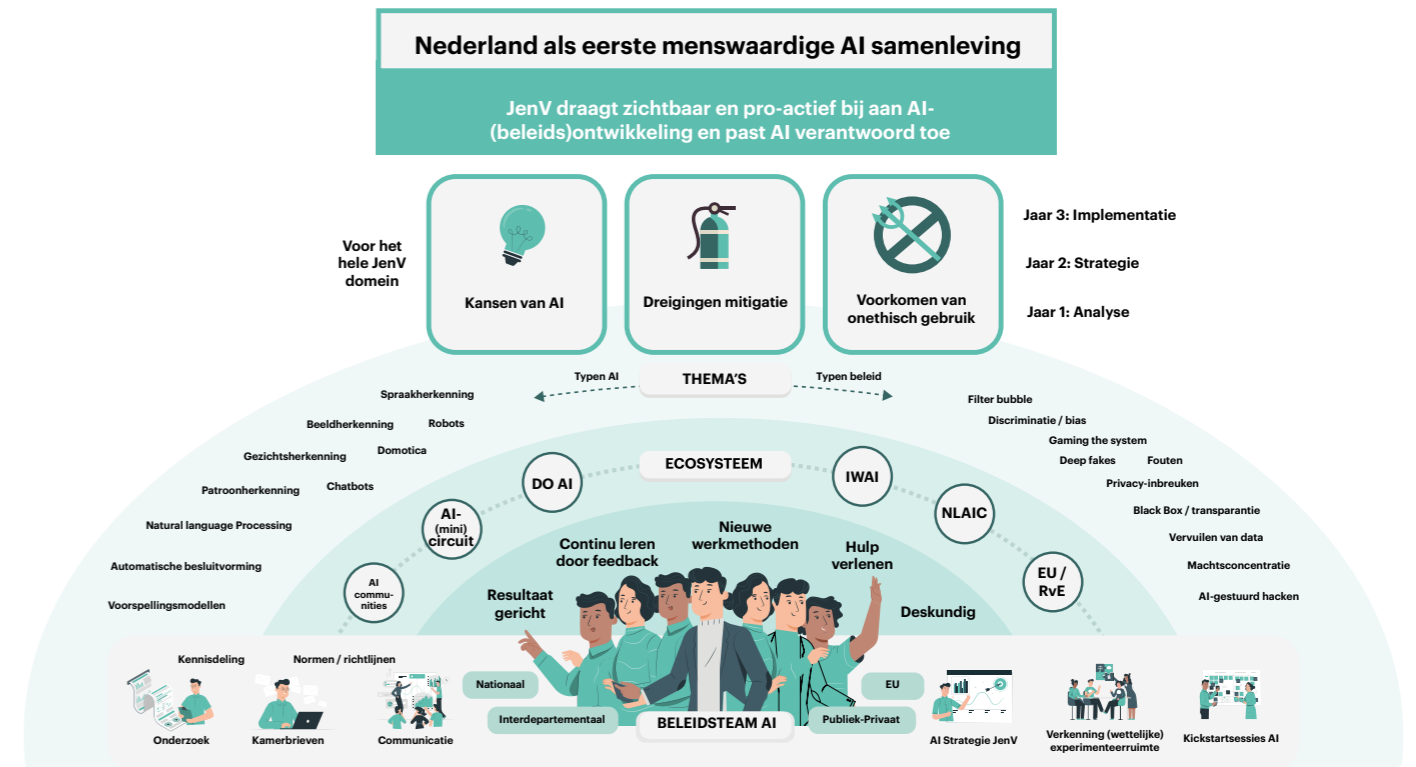


Figure 3: the role and activities of the AI team in relation to the mission

Figure 3 also resembles the topics, regarding the types of AI technologies and policies, the AI team is concerned with. These have been extracted from Table 1, which provides an overview of the application possibilities, technologies and threats which are relevant for the ministry. These lists are retrieved from the Quartermasters plan on AI (2020).

Table 1: three lists that consider application areas, technologies and threats for the ministry.

Possible application areas that could benefit the use of AI:	AI technologies that can be distinguished:	The threats caused by AI relevant for the ministry are:
<ul style="list-style-type: none"> minimize victimization accelerate justice to implement punishments more effectively increase chances fight fraud effective business operations prevent disruption tackling undermining ethical research 	<ul style="list-style-type: none"> image recognition voice recognition pattern recognition natural language processing chatbots automatic decision-making prediction models person recognition home automation 	<ul style="list-style-type: none"> deep fakes black box (lack of transparency) mistakes gaming the system AI-driven hacking Power concentration of Big Tech Filter bubble Privacy invasion Data pollution Discrimination / bias

1.2 Project Approach

In this section the project aim, research questions and research approach is discussed.

1.2.1 Project aim

The aim of this research is to provide a structural approach for policymakers to enable the formulation of policies for regulating new technology. The common challenges in the collaboration of policymakers and executing parties regarding AI experiments will be analyzed through several use-cases. The chosen use-cases focus on AI technology which is used to help police teams in detecting criminal activity. A framework illustrated through supporting tools for the AI team will be proposed that facilitates human-centered AI development. The solution should comply with the EU commission ethical regulations and enable mutual learning for the involved parties.

The ministry stated the main challenge to be:

“How can AI experiments enable the formulation of policies to steer human-centered AI adoption?”

1.2.2 Research questions

To answer the challenge, I have split the explorative study into 5 research questions (RQ):

1. What is the added value of AI experiments for the ministry?
2. What are the challenges in the current approach to regulate technology?
3. What are the challenges in the current AI experiments?
4. What are the necessary building blocks to enable organizational learning?
5. What can the AI team offer policymakers to formulate policies on the development of human-centered AI applications?

RQ 1-3 will be answered in Chapter 2: Explorative study: describing the insights gained from a deepdive in AI, the political context, and the current state of AI experimentation, and Chapter 3: Problem Reframing

RQ 4 will be answered in Chapter 4: Design Vision presenting the building blocks in a manifesto.

RQ 5 will be answered in Chapter 5-7: an actionable toolkit based on the framework, including the evaluation and further advice

1.2.3 Process approach

‘Artificial Intelligence’ and ‘policymaking’ are overarching words that consider a large field of research and are complex to grasp. For this research, additional parties were involved next to the TU Delft supervisory team, the company mentor, and the colleagues of the AI team. Consulting policymakers, ethics and AI consultants informed the research for a greater understanding of the context. AI startups which took part in the SBIR experiments supported the empirical research. Experts covering the fields of policymaking, social design, journalism, IT, safety enforcement, ethics and philosophy of technology enriched the multiple perspectives on the design as well as supported the validation. For the list of events and interviews, see Appendix A. During this thesis the 5-step Design Thinking method is used to structure the research approach, represented in Figure 4.

DESIGN THINKING: A NON LINEAR APPROACH

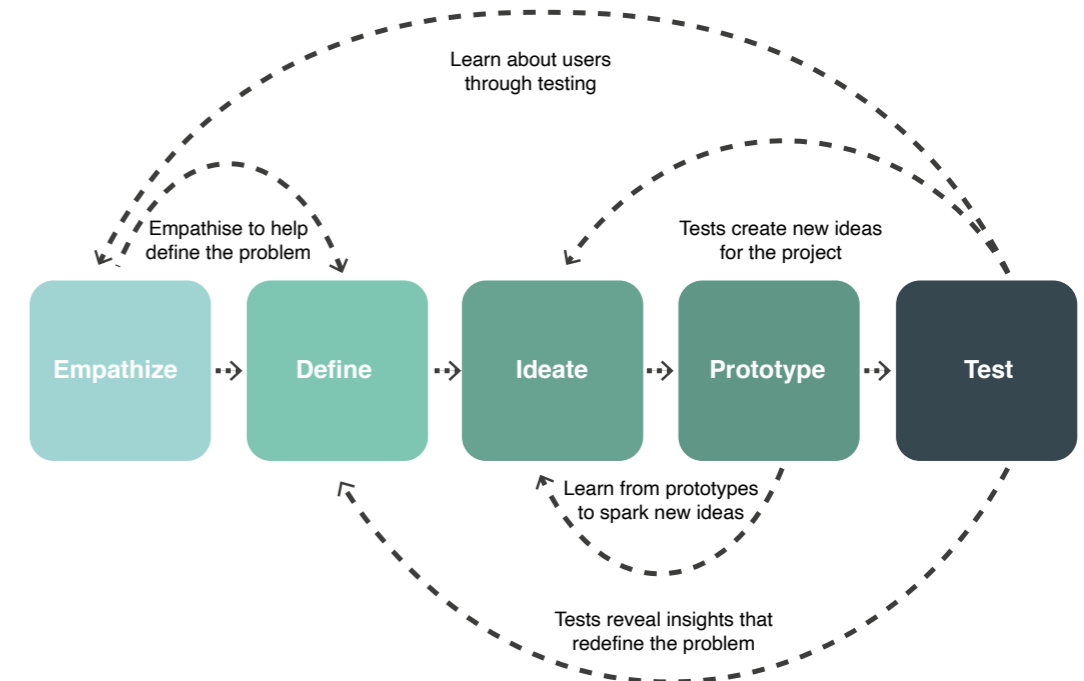


Figure 4: the 5 step model of Design Thinking

Empathise

The process of empathizing considers deep qualitative research techniques, presented in Figure 5. As the problem was stated as an explorative study, the data is collected by interviews, observations as well as generative sessions. This process required a close collaboration with the AI team, thus being actively involved in the daily team practices.

Interviews: open and semi-structured interviews were held to retrieve explicit knowledge on the political environment, the organizational structure and processes.

Observations: observing the dialogue of the weekly meetings and joining relevant webinars hosted by the ministry revealed insights in the patterns of overlapping observable needs. Proactively taking part in the team practices also led to a richer understanding and valuable insights, a network of resources and references to relevant literature.

Generative Sessions: organizing sessions helped to understand ideas, dreams and latent needs. It revealed the value of AI, the desire for a normative approach, the dream to feel in control, and the vision of how the toolkit could be implemented. Several elements in this thesis consisted of research through design.

DATA COLLECTION TECHNIQUES

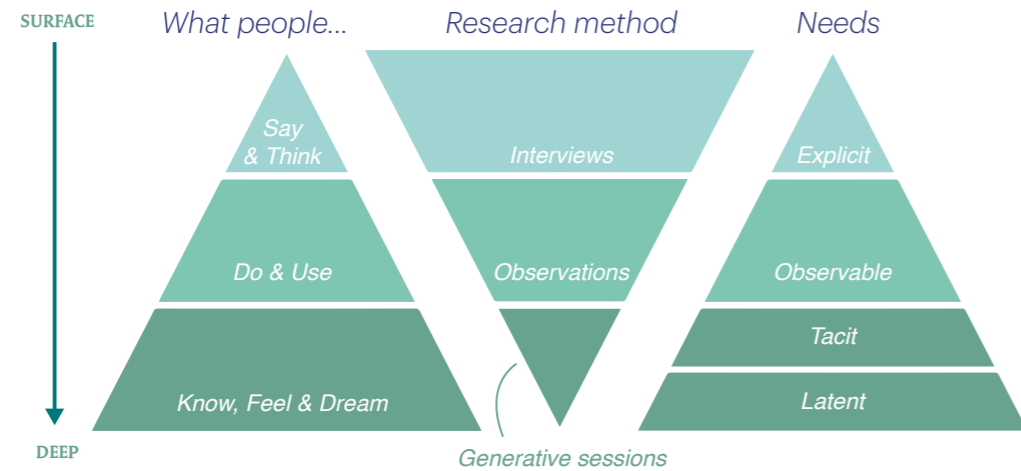


Figure 5: the qualitative research techniques for data collection (Sanders & Stappers, 2012)

Define

Multiple interviews with the AI team, closely related board members and colleagues within the department provided me with a better understanding of the organization, the process of policymaking and their challenges in relation to the rapid pace of technology development. Facilitating a weekly online Design Thinking course at the start of this research provided clarity on the perceived role, the challenges and thinking patterns of the team. Interviewing AI startups with a generative tool gave me insight to the current governmental experimentation process and the collaboration between the public and private sector, which also provided an external perspective on the topic. Creating trend cards and hosting a generative session with students helped to shape an overview of AI developments in other industries in order to define the added value of AI. Clustering the identified problems (from unfamiliarity) revealed the complexity and interconnectedness representing a 'wicked problem'. Literature on how to approach wicked problems helped the emergence of a framework for agile policymaking and a vision for the ministry to adopt this process innovation.

Ideate & Prototype

Translating agile principles to policymaking resulted in a manifesto with the acronym 'CERTIFIED', representing the needed mindset for this transition management. Design requirements were extracted from the problem clusters and inspired by the manifesto. An online brainstorm was held to fuel the inspiration for the ideation. A toolkit including a card deck and canvases were created to facilitate agile policy development. In particular, the canvases shape the formulation of a Minimum Viable Policy, which is a tangible and actionable perspective to execute the 'CERTIFIED' approach. The conceptualization of the toolkit followed an iterative process of multiple testing rounds, validation interviews and adjusting the toolkit elements based on the observation and evaluation.

Test

A final pilot was held to represent a real-life implementation experience with the valid panel of stakeholder representatives. As part of the implementation plan, the conducted research consisted of an inclusive approach which generated attention and ownership of the acknowledged challenge, the manifesto and the toolkit. By actively involving a wide range of parties to interviews, validations, sparring moments and keeping them up to date, the level of traction could be monitored and validated. To evaluate the success of implementation, the toolkit was made public in a conference and close attention has been paid to the feedback on its desirability. Being invited to share insights on the research, explaining the designerly approach and joining panels and a conference as an inspiring keynote speaker provided the research a spotlight. Hence, a higher chance for successful implementation. With the aim to make 'CERTIFIED' a recognized acronym, it might continue to spread (like wildfire) through the organization.

1.2.4 Literature scope

In the literature research, the intersection of the design, AI and political context is explored. The literature scope is visualised in Figure 6. Within the design field, literature on human-centered design, design thinking, agile principles, transition and systemic design have been explored. A deepdive in AI included distinctions in AI technologies, the added value of AI, the main challenges, ethical guidelines and inspection tools. Topics within the political domain that have been explored consider the complexity theory, legislation, human rights, multi stakeholder perspectives and the process of policymaking.

THE INTERSECTION AS SCOPE

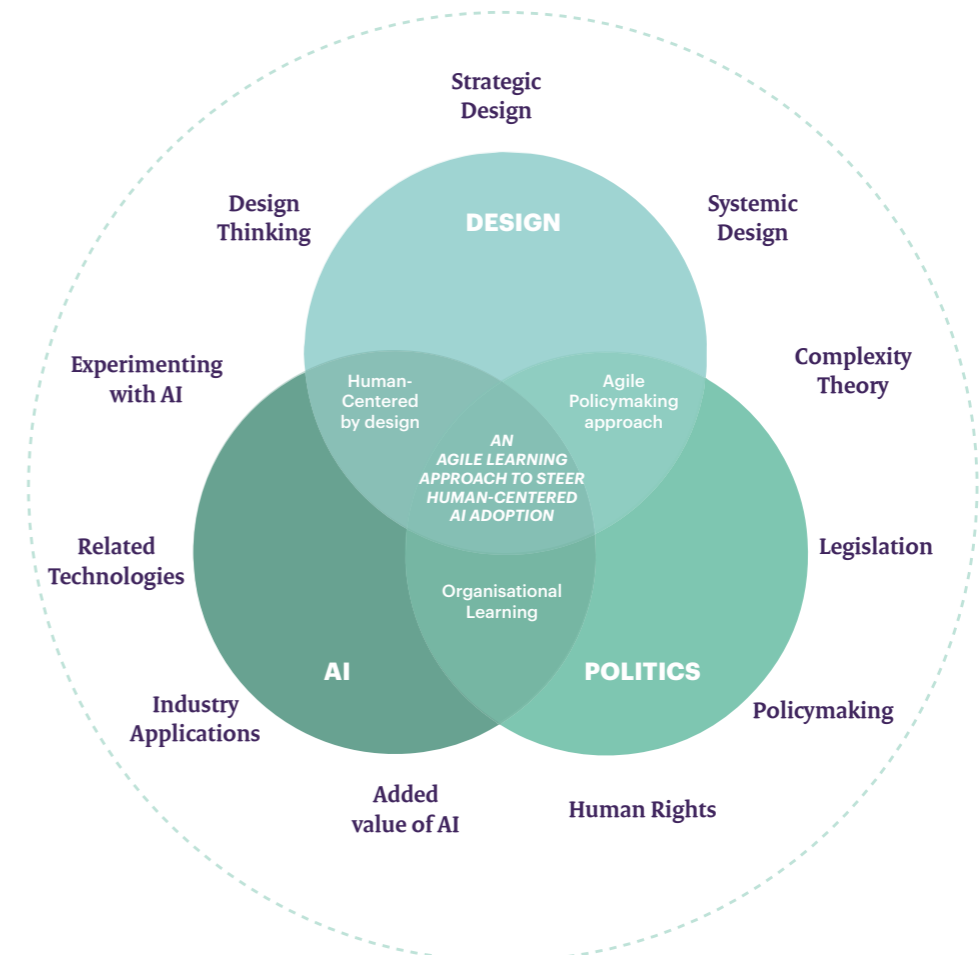


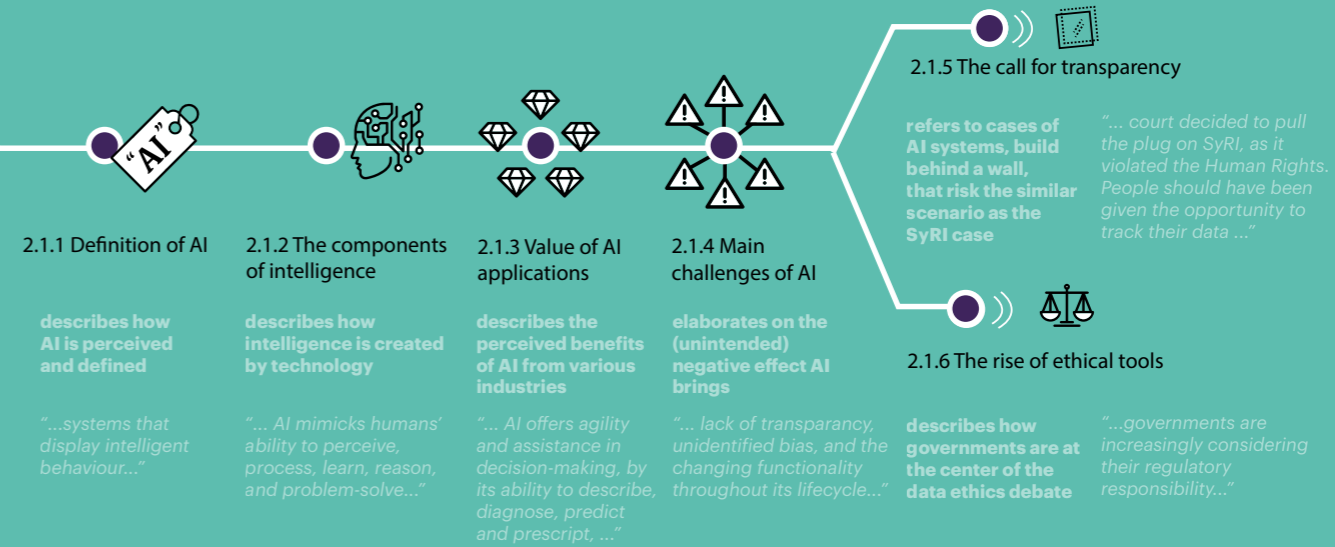
Figure 6: the intersection of the AI, design and political context

02 EXPLORATIVE STUDY

This chapter covers the explorative study in the field of AI, the political context of formulating policies, and the governmental experimentation approach. Each section ends with a summary of the gained insights and conclusions.

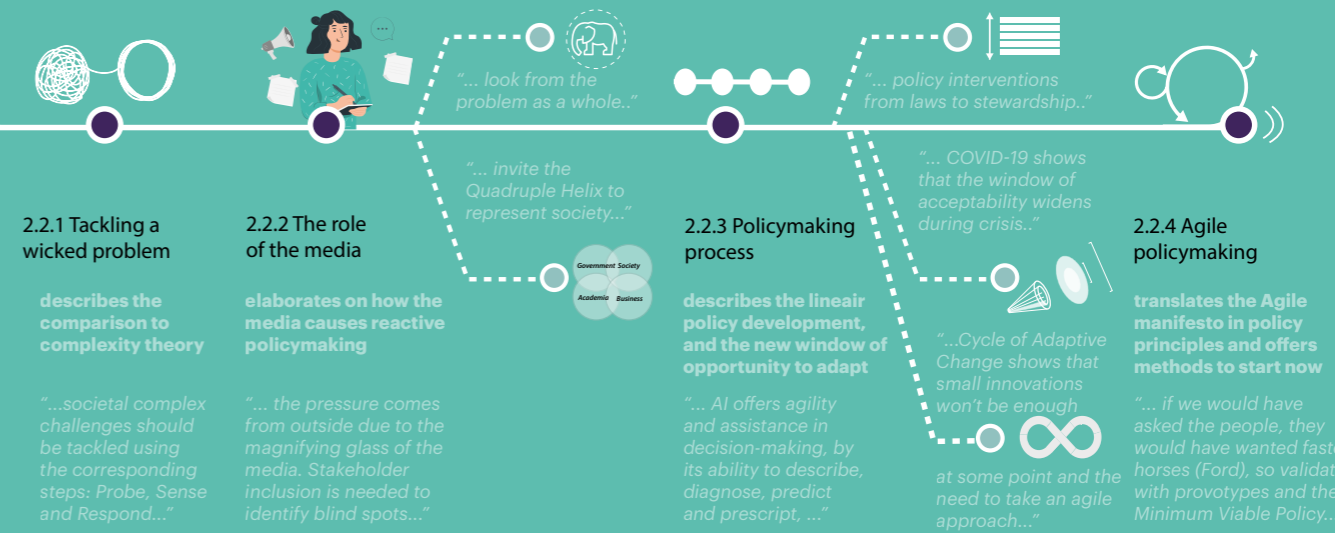
VISUAL READING GUIDE

2.1 DEEPDIVE IN AI



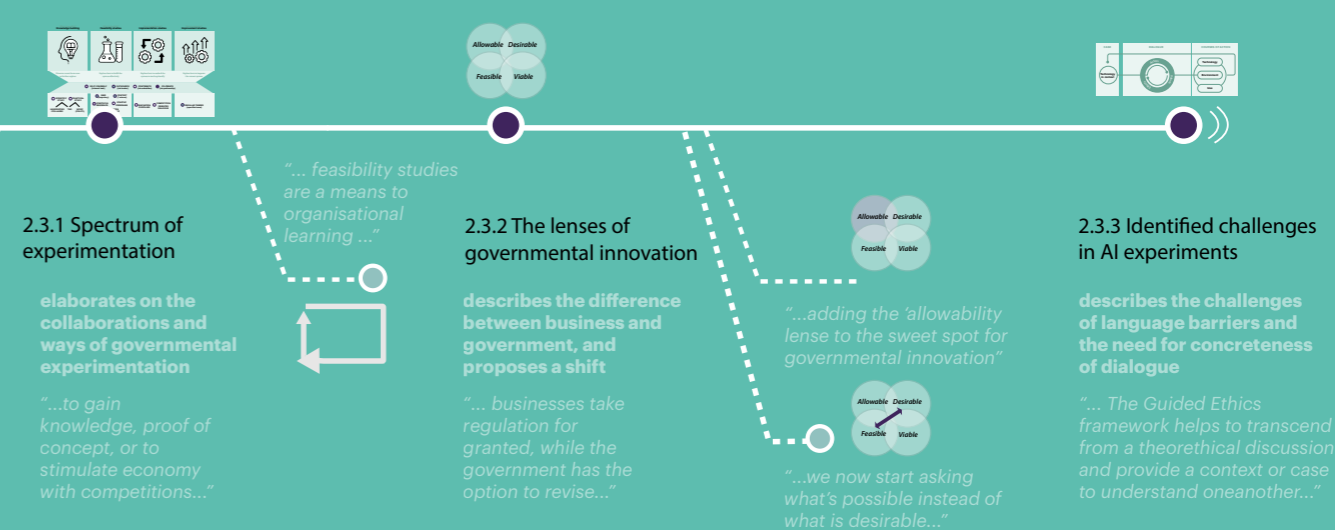
2.1.7. Insights & Conclusion

2.2 THE POLITICAL CONTEXT



2.2.5. Insights & Conclusion

2.3 CURRENT STATE OF AI EXPERIMENTATION



2.3.4. Insights & Conclusion

2.1 Deepdive in AI

In this section, the definition of AI, the components of this intelligence, and the added value for the ministry are discussed. This section also elaborates on the main challenges and the societal effect these applications entail. The section ends with the gained insights of this part in the format of 'building blocks'.

2.1.1 Definition of AI

There are multiple initiatives that try to grasp a definition of AI, however there is no generally accepted definition of AI that is consistently used by all stakeholders. A research on shared awareness when talking about AI (Elsevier, 2018) that discovered that different parts of society - media, education, science and industry - talk very differently about AI. To quote a few resources:

- *"AI is often described as a container concept that describes the ability of computers to mimic human judgment"* - Maria de Kleijn-Lloyd
- *"Artificial Intelligence is defined as applications that can sense, comprehend, act and learn leveraging machine learning, natural language processing and computer vision to automate and augment human activities."* - Accenture
- *"AI are intelligent systems that can independently perform tasks in complex environments and improve their own performance by learning from experiences."* - Nationale AI cursus

The ministry chose to adopt the definition stated by the European Commission on AI: *"AI refers to systems that display intelligent behavior by analyzing their environment and - with a degree of independence - take action to achieve specific goals."* To form a consistent picture and provide an understandable explanation of what AI is exactly, the ministry provided a free online AI course (app.ai-cursus.nl) to create awareness and explain what this technology entails.

2.1.2 The components of intelligence

Looking at the definition of AI to display intelligent behavior, we can distract the key components of intelligence from humans. A review on how humans create intelligence led to a distinction in 5 components: to perceive, to process, to learn, to reason, and to problem-solve. (Hao, 2018; Chui & Kamalnath, 2020) Combined with a set of moral values and the energy provided by a healthy lifestyle, they form the ingredients of human intelligence. This section ends with a visual representing the summarised insights (see Figure 7). For the extensive research on the 5 components, see Appendix B. In short, an explanation of the 5 components is described in this section.

To perceive: Humans collect information by perceiving their environment using their senses. As humans have senses, an AI distracts the data from sensors. Data could include for example: images, sound or voice recordings, a GPS location or temperature measurements. Where eyes represent vision, a camera can be combined with image processing or facial recognition. Computer vision can also analyse patterns to read text and respond with chatbots. Where ears represent sound, recordings can be combined with voice recognition software or Natural Language Processing to help transcribing. Where fingers represent touch, geo-location or movement sensors can be combined with smart robots to automatically pick up and move fragile objects. Where a tongue represents taste, heat sensors can detect and monitor a desired temperature.

To process: We distinguish four application areas of AI that contribute to knowledge building: to *describe, diagnose, predict, and prescript*.

To learn: Learning is an overarching component. Within the four application areas of AI, descriptive analysis doesn't necessarily include AI as there is no learning component in the analysis itself. When a prediction or prescription element is present, there is a decision element involved. In this case machine learning occurs. Machine learning can be distinguished in reinforcement learning, unsupervised learning and supervised learning. (Rozenblum et al., 2020) In case the system uses deep neural networks, we speak of a special type of machine learning: deep learning. These neural networks are interconnected layers of software-based calculators known as "neurons" which form a neural network. The network can ingest vast amounts of input data and process them through multiple layers that learn increasingly complex features of the data at each layer.

To reason: To reason includes drawing inferences appropriate to the situation. These can be classified as either deductive or inductive. Deductive inferences guarantee the truth of the conclusion (i.e. "X must be either A or B. He is not in A, therefore he is in B.") Inductive inferences guarantee truth of the premise that supports conclusion (without absolute assurance)(i.e. "Similar accidents of this sort were caused by A. Therefore this accident was caused by A.")

To problem-solve: Problem solving by AI may be characterized as a systematic search through a range of possible actions in order to reach some predefined goal or solution. We can distinguish weak (narrow) AI, by the special-purpose, from strong (general to super) AI, which has a general purpose. In this range, narrow AI is represented by an AI that is optimized to make coffee, but provides an error when the machine is out of coffee beans. General AI is when this machine is able to solve a variety of problems by for example asking the neighbours for the coffee beans. Super AI represents the capability to solve problems beyond human capacity (i.e. writing a best-seller novel in the time you take a sip of your fresh coffee)

AI MIMICKING HUMAN INTELLIGENCE

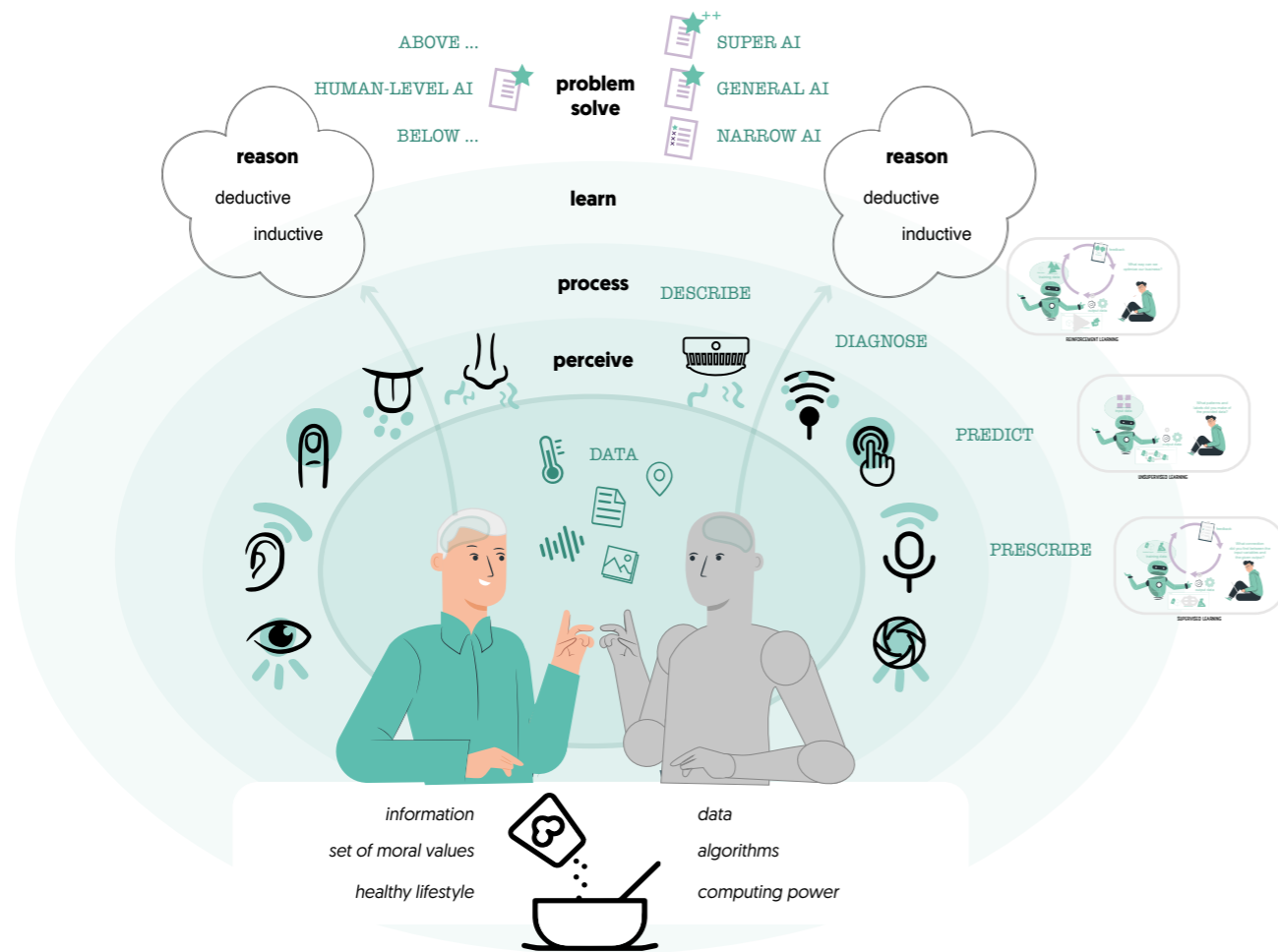


Figure 7: the summarized insights on the approach to mimic human intelligent behaviour with AI.

2.1.3 Value of AI applications

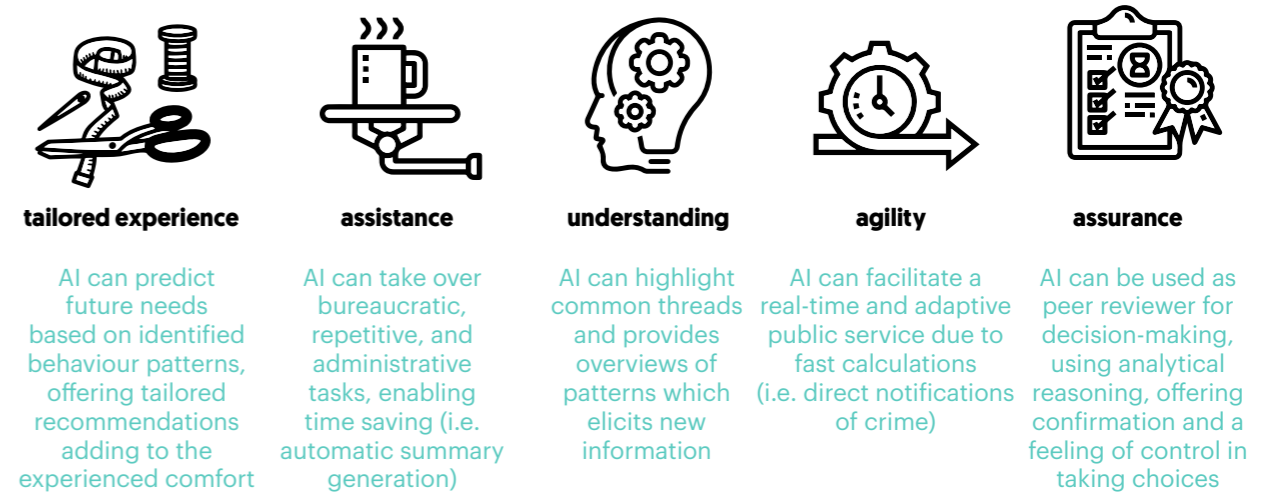
A set of cards were created of different domains to explore the added value of AI for the ministry. During a generative session the value of AI was researched by reflecting on the cards representing current AI applications in the different industries. Figure 8 provides a glimpse of the facilitated session. These industries included: healthcare, media, construction, finance, entertainment, and public services. The used cards can be found in Appendix C.

RETRIEVING THE VALUE OF AI



Figure 8: the generative session to retrieve the value of AI

The added value of AI for the ministry covers the following topics:



During the session, the participants mentioned that the value is prudent to how the data is handled, as the application could also be hacked, misused or data could be reused which would raise the ethical discussion. One of the participants mentioned *“The AI applications can only add value when it weighs out the negative consequences.”* The exploration session quickly transformed in a discussion about data privacy, data monopoly, governance, responsibility and transparency. Questions such as *‘Where is the line?’*, *‘Who is in power?’*, *‘Who has access to my data?’*, *‘Who to blame when the system causes mistakes?’*, *‘What personal data is considered and what is it used for?’*, and *‘How does the AI draw conclusions?’* showed the need for clarification when we talk about the use of artificial intelligence

It also became clear that the positive or negative stance towards AI differs depending on the framed organization’s intent to improve their service using machine learning. When machine learning is used to improve someone’s health, we find the use of AI more acceptable., whereas a private organization using AI to make them more profitable, resistance arises questioning what’s in it for them? When the police use AI to fight crime, the question is *‘what do you need from me?’* whereas in the case the court uses AI to support trails, the question is *‘what is my saying in this?’* Whose statement do you value more? Reflecting on this phenomena, when AI is used for services that recognizably benefits the individual, the less resistance is perceived. When someone is sick, your consciousness awakens that you can’t take health for granted. The same counts for safety on the streets where we expect the government to take care of our safety and to be held responsible.

2.1.4 Main challenges of AI

In recent times, several books have been published that speculate the utopian and dystopian views on the future of AI. Some disturbing but also hopeful stands are presented about the connection between humans and AI. Where the Netflix documentary *‘The Social Dilemma’* addresses the influence of AI nudging humans unconsciously as slaves in their own destructive business models. *Life 3.0* by Max Tegmark features the most important conversation of our time: *‘How to be human in the age of AI?’* As Verbeek (2020) mentions *‘it’s too late to get rid of AI’*, and the amount of AI applications continues to grow, what will that future look like? Will superhuman intelligence be our slave, or become our god? How can we grow our

prosperity through automation, without leaving people lacking income or purpose? How can we ensure that future AI systems do what we want without crashing, malfunctioning or getting hacked? Should we fear an arms race in lethal autonomous weapons? Will AI help life flourish as never before, or will machines eventually outsmart us at all tasks, and even, perhaps, replace us altogether? "AI is a rare case where we need to be proactive about regulation instead of reactive. Because I think by the time we are reactive in AI regulation, it's too late." (Musk) The SAPAI report, as well as the EU Commission whitepaper, the report 'Menselijke grip op algoritmen' by Middendorp (2019) and the research by Hooghiemstra (2020) stress the need for a human-centered approach.

The EU Commission's whitepaper introduces the guidelines for achieving excellence and trustworthy AI. It presents the seven aspects that need to be considered, represented in Figure 9:

GUIDELINES FOR EXCELLENCE AND TRUSTWORTHY AI

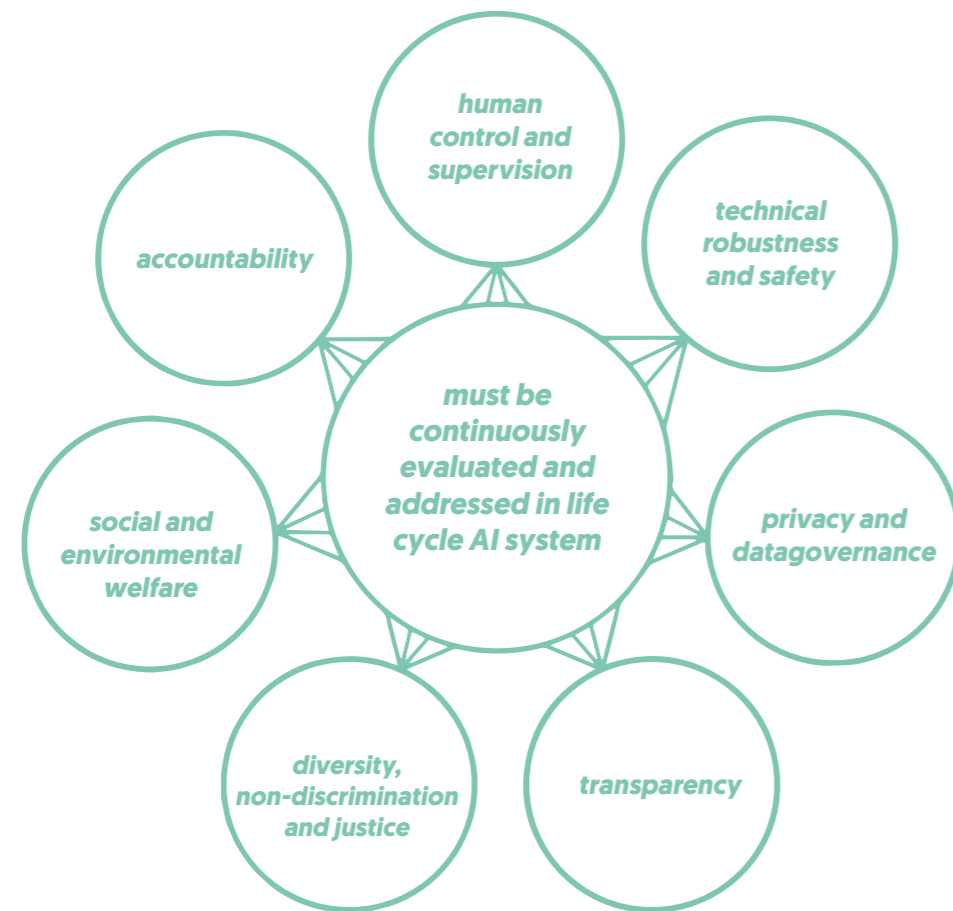


Figure 9: the guidelines for excellence and trustworthy AI (EU Commission, 2020)

As these guidelines should be continually evaluated and addressed in the AI system's life cycle, the data mining lifecycle, adopted from the CRISP-DM methodology, is used to plot the identified challenges. The CRISP-DM methodology stands for Cross Industry Standard Process for Data Mining, that describes commonly used approaches that data mining experts use to tackle problems in traditional data mining. (Shearer, 2000) An overview of the challenges is presented in Figure 10.

CHALLENGES WITHIN THE AI LIFECYCLE

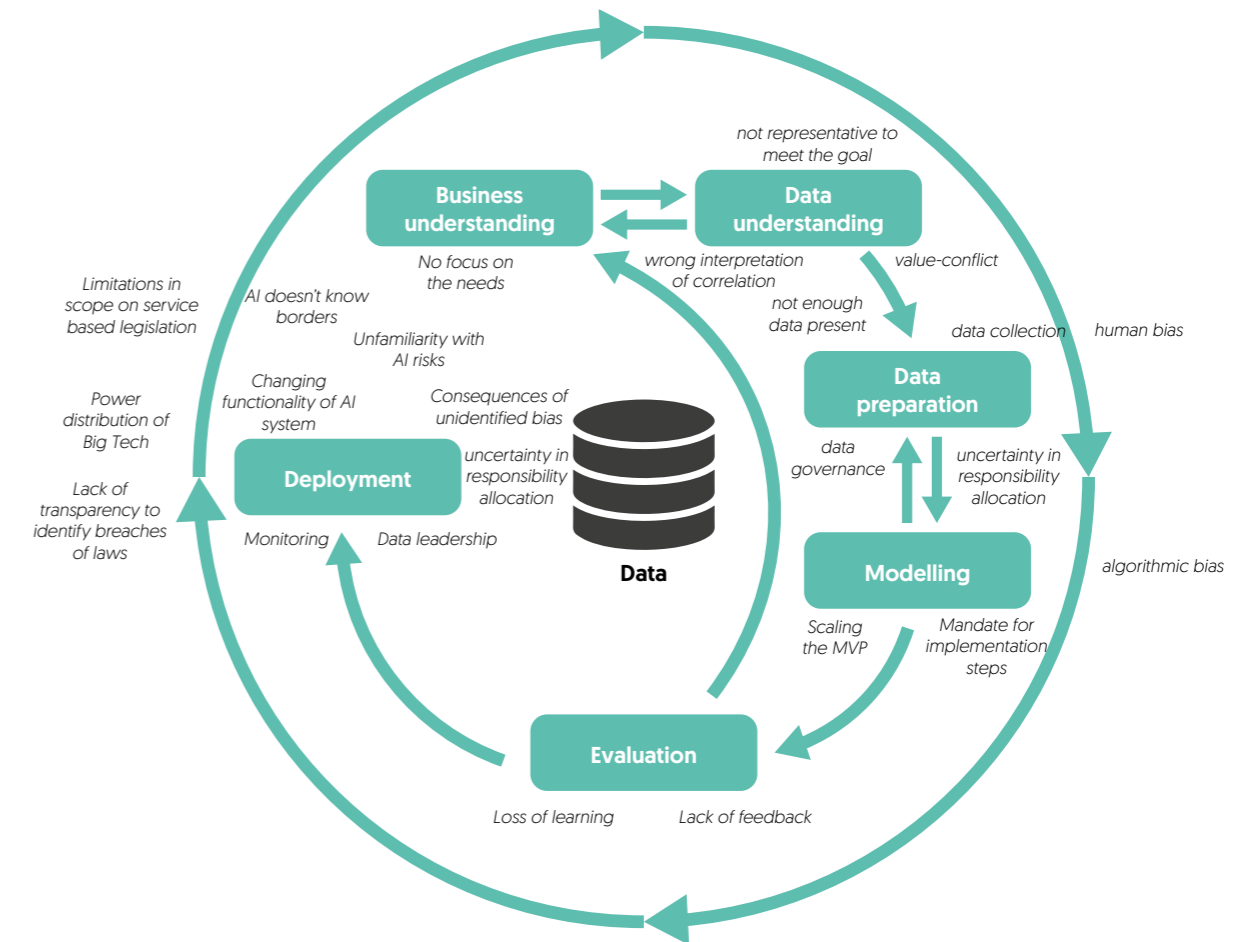


Figure 10: the identified challenges throughout the lifecycle of an AI system plotted on the CRISP-DM lifecycle

The challenges of AI posed by literature have been clustered, which resulted in the following list of overarching challenges:



Limitations on the scope of service based legislation:

Current legislation applies to products, not to services nor service-based technology. Lack of AI transparency to identify breaches of laws. AI operates from a black box, making it harder to identify where in the system an unintended act of violation occurs.



Changing functionality of AI systems:

AI has an adaptive character by nature, indicating that AI may be in line with legislation at one point in time, but over time it might violate the law over its lifecycle. It is difficult to be aware of that turningpoint and existing legislation predominantly focuses on safety risks present at the time of placing the 'product' on the market.



Uncertainty in responsibility allocation:

Who owns the weather? It appears that such complex phenomena aren't felt as an individuals' responsibility.



The power distribution of big tech:

IBM, Microsoft and Amazon have decided to halt the sales of Facial Recognition technology. The shift towards Big Tech creating their own guidelines before the government announces also shifts the power perception. The government lacks skills in data science, thus relying on strategic choices of Big Tech, providing them a powerful voice.



Unfamiliarity with AI risks:

Unidentified bias could lead to discrimination by either the bias in the data which is fed to the system, or the way the algorithm is designed.

- Human bias - These algorithms are created by people, who all have biases. They are never fully "objective"; rather they reflect the world view of those who build them and the data they're fed.
- Algorithmic bias - Machine learning algorithms generally rely on correlation, rather than causal relationships. Correlation tests for a relationship between two variables. However, seeing two variables moving together does not necessarily mean we know whether one variable causes the other to occur.



High risk with scalability of AI:

There will always be a percentage of false positives. The scale to which the AI system is deployed will therefore be crucial. It's not inevitable that humans make mistakes, although by human bias in algorithms, the scale the algorithm is operating in can make it disastrous. In the case of a high-risk algorithm, meaning it takes decisions that are considered as highly impactful, the bigger the scale it is deployed, the higher the percentage of false positives and false negatives, the more damage it will bring.



AI is not comply to national borders:

As the system runs on data it is fed, and data can be extracted from the world wide web, it is not aware of national borders. Restrictions to distribute facial recognition software in Belgium while it is widely embedded in the surveillance regime of China, makes one question what rules apply when AI systems from other countries collect your data online.

2.1.5 The call for transparency

Cameras, algorithms and infinite flows of data. More and more cities are using advanced techniques to make life in the city safer and more efficient. SomIt may sound useful and attractive, but things don't always go well, as it turned out in Rotterdam in 2020. A scan car was deployed in a park to check whether people kept the 1,5m distance. A far too heavy measurement, opponents thought. They think smart cameras are not suitable for these types of situations. The discussion about privacy versus security has been raging for some time. But now that technology can no longer be ignored in our lives, the question is how we can use technologies in such a way that our cities become smarter. Should we want that at all? And where is the limit? Research by Kantar Public (2019) showed that citizens attach the utmost importance to the protection of privacy, followed by the drafting of laws and regulations. In third place is the need for knowledge about how AI works and how to deal with it. 7% of the respondents want more insight into AI and the consequences of AI (think of an explanation of algorithms).

One of the major topics related to the use of AI is the call for transparency. To stress the importance of transparency in the operation and application of algorithms, a reference is made to the SyRI case (see Figure 11). The System Risk Indication system (SyRI) has been used as a means of enforcement, to uncover fraud cases by creating risk profiles of potential fraudulent people to be thoroughly investigated. The system linked all kinds of information about citizens, from, for example, the Tax Authorities, the UWV, the Social Insurance Bank.

SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens

Den Haag, 05 februari 2020

De rechtbank heeft vandaag uitspraak gedaan in een zaak over het Systeem Risico Indicatie (SyRI). SyRI is een wettelijk instrument dat de overheid gebruikt voor de bestrijding van fraude op bijvoorbeeld het terrein van uitkeringen, toeslagen en belastingen. De rechtbank is van oordeel dat de wetgeving die de inzet van SyRI regelt in strijd is met hoger recht. De wetgeving voldoet volgens de rechtbank niet aan artikel 8 van het Europees Verdrag voor de Rechten voor de Mens (EVRM). Dit artikel beschermt het recht op respect voor het privéleven.



Figure 11: snapshot of SyRI newsletter article

The judge concluded that the SyRI system violated Article 8: the right to private life, regarding the invasion of privacy. According to the court, it was also difficult to see how a data subject could defend himself against the fact that a risk report has been made with regard to the person in question.

The right to respect someone's privacy includes that a data subject must be given a reasonable opportunity to track his or her data. The importance of transparency, with a view to verifiability, is also important because the use of the risk model and the analysis carried out in that context entails the risk that (unintended) discrimination will occur. So, the way in which the system worked could not be checked, which made the algorithm a black box.

As this system was not in line with the European Convention on Human Rights, the government stopped using SyRI. This case showed the conflict, where the rights to respect privacy weights out the attempt to bring justice by uncovering fraud cases.

The aim of these guidelines is to increase the transparency (in both explainability and technical transparency) of algorithms, its operation and its application. Transparency around algorithmic data analysis can contribute to the trust that citizens have in these analyzes. To that extent, transparency is primarily a means for the government, where trust by citizens is the goal. Transparency will also improve the ability of citizens to check the government's working methods in data analysis and thus contribute to the most balanced relationship between citizen and government. Transparency can also lead to better compliance with legislation and regulations. Presenting variables or threshold values that the government uses in data analyzes can on the one hand promote calculating behavior, but on the other hand it can be used as a means to prevent certain actions by citizens (function of nudging).

2.1.6 The rise of ethical tools

The call for transparency describes the need to communicate how political decisions in policy issues are being made. This need also entails gaining insight in the ethical reasoning behind the decisions. However, the ethical dimension of policy issues is not always immediately recognized.

In essence, the government is there to help society preserve or realize public values. Public values are values that are widely supported and of which at least a majority of the population believes that they should be guaranteed by the government or that are difficult or impossible to achieve without government. Examples of public values which are relevant for this thesis are national security, safety and justice.

Primarily, citizens and companies are responsible for the preservation or realization of their own security, safety and justice. However, where this is not possible due to overarching conflicting values, an appeal is made to the government. If the issue belongs in the public domain, the government will look for a solution to the problem together with the parties involved. This is not always easy, especially when public values clash.

The continuous developments in our environment are changing the public's perceptions of the meaning and weight of public values. Public values do not disappear, but the mutual ranking and the extent to which we consider them important do change over time. As AI systems include certain ethical decision making, the question of whether the system can be declared as 'ethical' has become a prominent topic of conversation. *"The conversation has shifted from you can't do it ... to you shouldn't do it! That shift has many parallels in other domains, from vaccination, to flight, to splitting the atom, to gene manipulation."* (Ford, 2015) In addition, the dimension of time has recently become an urgent issue to address.

"Self-learning algorithms might change in a way that they do not meet the criteria of the ethics assessment anymore, or a data model reflecting the political values of the municipal council requires changing after the next election." (Van den Hoven, 2020)

"As technology tracks huge amounts of personal data, data ethics can be tricky, with very little covered by existing law. Governments are at the center of the data ethics debate in two important ways." (Deloitte, 2019)

- The government "owns" a massive amount of data about citizens, from health records to what books a citizen checked out of the library.
- The government is a "regulator" of the corporate use of data collected online.

Governments are increasingly considering their regulatory responsibility. For instance, the European Union's General Data Protection Regulation (GDPR) provides strict controls over cross-border data transmissions, gives citizens the right to be "forgotten," and mandates that organizations, including government agencies, provide "data protection by design" and "data protection by default."

Tallacchini (2009) explains the need for research on the normative instrument of 'ethics'. *"The institutionalization of ethics and the direct influence of politics on how ethics bodies frame their opinions have been widely recognized and explored in the last few years. Less attention has been paid to what kind of normative instrument "ethics" as an institutional phenomenon has become in the rule of law, and which institutional powers it has depended on."* Therefore, a variety of ethical tools have been explored to understand the pro's and con's of utilizing them. The preconditions for (Big) Data and AI cover the information security, data governance, data privacy, and public values and human rights. The ministry offers multiple tools that help meet these preconditions. A few tools are listed and explored for its use, the topics they cover and the possible improvements (see Figure 12)

- BIO: information security
- DAMA DMBOK: data governance
- PIA: Privacy Impact Assessment
- DEDA: Ethics and bias
- Transparencylab: transparent algorithms

ANALYSIS OF CURRENT AVAILABLE ETHICAL TOOLS

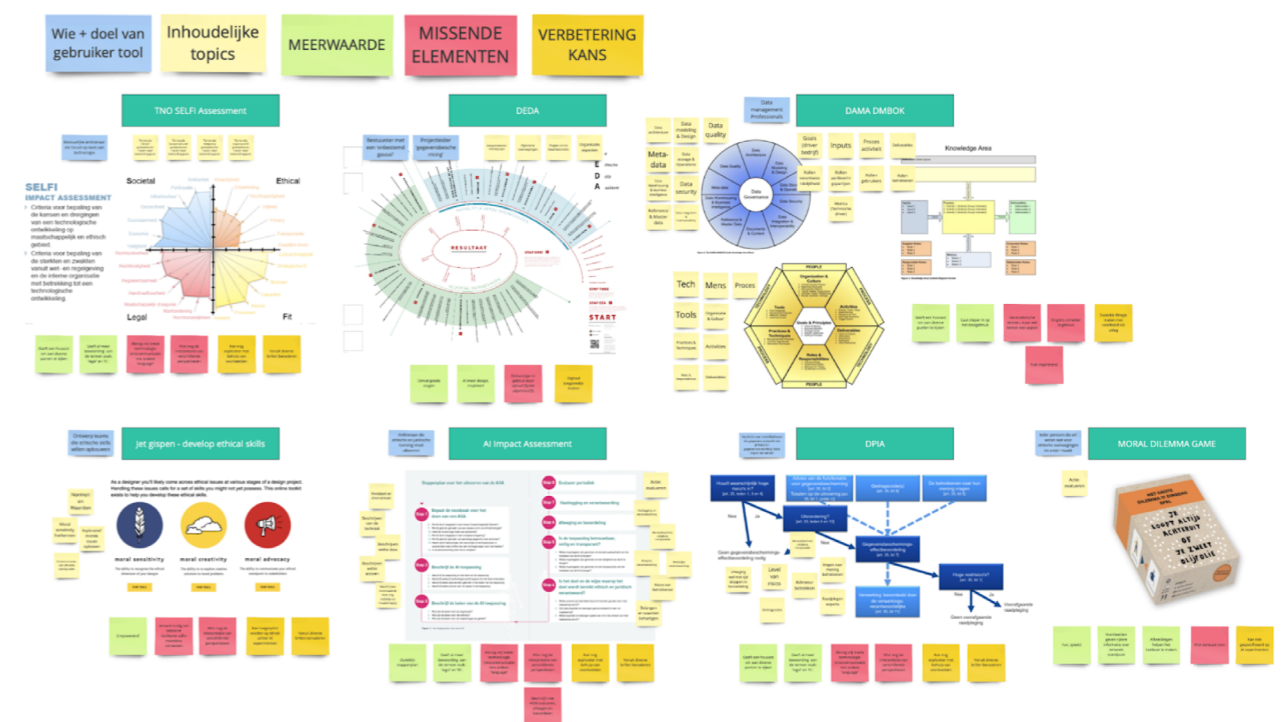


Figure 12: exploring the range of ethical tools for its use and improvements

Combining this study with the conducted interviews, it came to light that these tools are not perceived attractive in use. The tools usually cover over 50 questions or are experienced by the interviewed developing parties as a formality rather than a necessity to define the ethical considerations to undertake. Hence, the current ethical tools used by the ministry could be improved making the interaction more attractive. Interviewees stated that a 'fun' element would increase their willingness to not skip this formality. As these tools are considered crucial, of which conducting the Privacy Impact Assessment (PIA) has become a mandatory step, it should 'invite for play'.

It is often said that people like being in control of things (Alloy & Abramson, 1979; Vohs & Baumeister, 1975) However, the only thing we really control is the process through which we come to our decisions. To provide a structure to this process is the only rational way to handle the unknown. The ability to clearly state the known unknowns helps to gain the feeling of control in the situation. All the above explains the rise of these ethical tools.

2.1.7. Insights & Conclusion

THE BUILDING BLOCKS



Creative:

Attractive ethical tools should 'invite for play'. We need to come up with creative problem solving ways to regulate technology. We should be open to exploration, questioning the current way of working. In order to come up with innovative solutions, we should look beyond what is possible within the regulatory frame and combine ideas and minds of different industries, like taking inspiration from healthcare measures or design principles. Designing for conflicting values lays a completely different basis for innovation, resulting in for example privacy-by-design.



Transparent:

The call for transparency is identified. People want to know what is happening with their data, how algorithms take decisions. It also shows that the government needs to tear down its wall and communicate what experiments happen, so society can formulate an opinion on the decisions behind the wall. This offers the possibility to provide feedback early in the process, leading to less often the need to pull the plug out of experiments.

2.2 The political context

In this section, the challenge of wicked problems, the effect of (the magnifying glass of) media, and the policy process are elaborated. It will also shed light on the learnings from the Covid-19 pandemic and the added value of shifting towards agile policymaking. The section ends with the gained insights of this part in the format of 'building blocks'.

2.2.1 Tackling a wicked problem

Characteristics of a wicked problem

Complex problems in policymaking are common (Holierhoek & Price, 2019). In 1973, Rittel and Webber attempted to understand the nature of these kinds of problems, assigning the title of 'a wicked problem'. Wicked problems are characterised as, "a class of social system problems which are ill-formulated, where the information is confusing, and where there are many clients and decision makers with conflicting values" (West Churchman, 1967, p. B-141). In more recent research, wicked problems have been described as problems with high levels of complexity, uncertainty, and divergence of values between different stakeholders (Head, 2008). In isolation, these factors do not make a problem wicked. However, once present simultaneously, they reinforce one another (ibid).

Although Rittel and Webber framed the concept in terms of social policy and planning, Conklin (2006) stresses that wicked problems occur in any domain involving stakeholders with differing perspectives. Recognising this, Rittel and Kunz developed a technique called Issue-Based Information System (IBIS), which facilitates documentation of the rationale behind a group decision in an objective manner (Rittel & Kunz, 1970). These visualizations of nodes between elements are often called "issue maps", used as a means of widening the coverage of a problem. By encouraging a greater degree of participation, particularly in the earlier phases of the process, the designer is increasing the opportunity that difficulties of his proposed solution, unseen by him, will be discovered by others.

A recurring theme in research and industry literature is the connection between wicked problems and design (Stolterman, 2008; Rittel, 1988). Design problems are typically wicked because they are often ill-defined (no prescribed way forward), involve stakeholders with different perspectives, and have no "right" or "optimal" solution. (Conklin, 2016) Thus wicked problems cannot be solved by the application of standard (or known) methods; they demand creative solutions. (Conklin, Basadur & VanPatter, 2007)

The finite problems tackled by science and engineering are relatively 'tame' or 'benign' – in the sense that the elements of a mathematics problem are definable and solutions are verifiable. By contrast, Head (2008) states that modern social problems are seen as 'ill-defined', interlinked, and relying on political judgments rather than scientific certitudes. In this sense, most major public policy problems are 'wicked' (Rittel and Webber, 1973), i.e. they are inherently resistant to a clear statement of the problem and resistant to a clear and agreed solution.

Dorst (2015) stresses that when organizations apply old methods of problem-solving to new kinds of problems, they may accomplish only temporary fixes or some ineffectual tinkering around the edges. Today's problems are a new breed -- open, complex, dynamic, and networked -- and require a radically different response. (Dorst, 2015)

Rittel and Webber (1973) identified 10 primary characteristics of wicked problems (see Figure 13):

1. There is no definitive formulation of a wicked problem, i.e. even the definition and scope of the problem is contested;
2. Wicked problems have no 'stopping rule', i.e. no definitive solution.
3. Solutions to wicked problems are not true-or-false, but good-or-bad in the eyes of stakeholders.
4. There is no immediate and no ultimate test of a solution to a wicked problem.
5. Every (attempted) solution to a wicked problem is a 'one-shot operation'; the results cannot be readily undone, and there is no opportunity to learn by trial-and-error.
6. Wicked problems do not have a clear set of potential solutions, nor is there a well-described set of permissible operations to be incorporated into the plan.
7. Every wicked problem is essentially unique.
8. Every wicked problem can be considered to be a symptom of another problem.
9. The existence of a discrepancy representing a wicked problem can be explained in numerous ways.
10. The planner has no 'right to be wrong', i.e. there is no public tolerance of initiatives or experiments that fail.

CHARACTERISTICS OF A WICKED PROBLEM

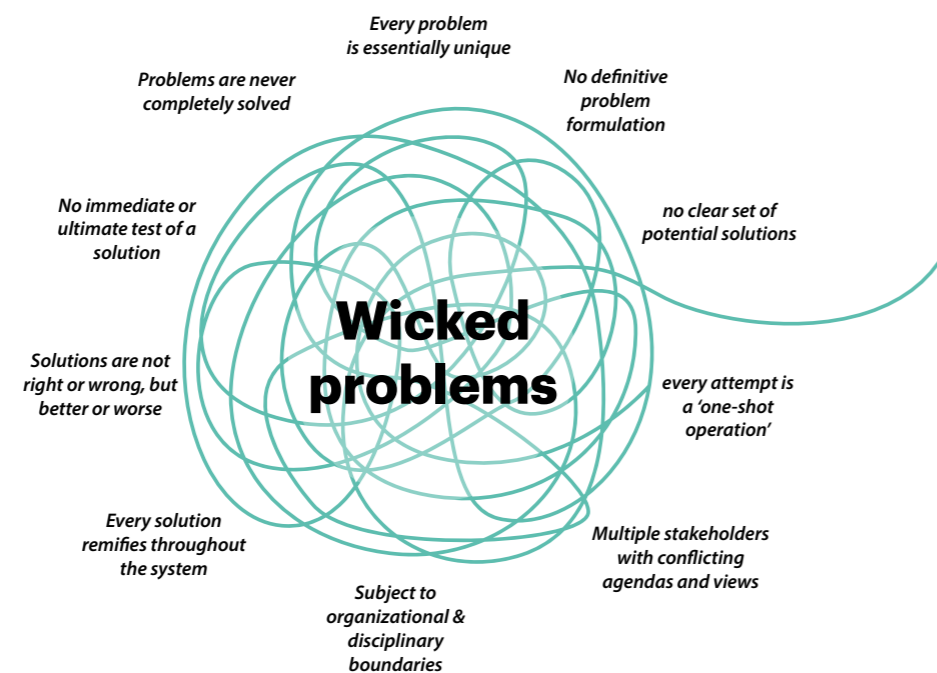


Figure 13: the 10 identified factors of a wicked problem by Rittel & Webber (1973)

The level of "wickedness" is measured by rating low, moderate or high on the three dimensions of:

- complexity of elements, subsystems and interdependencies
- uncertainty in relation to risks, consequences of action, and changing patterns
- divergence and fragmentation in viewpoints, values, and/or strategic intentions

A way to approach wicked problems

Snowden & Boone (2007) developed the Cynefin framework to broaden the traditional approach to leadership and decision making and form a new perspective based on complexity science, depicted in Figure 14. The framework sorts the issues facing leaders into five contexts defined by the nature of the relationship between cause and effect. Four of these (simple, complicated, complex, and chaotic) require leaders to diagnose situations and to act in contextually appropriate ways. The fifth (disorder) applies when it is unclear which of the other four contexts is predominant. Using the Cynefin framework can help executives sense which context they are in so that they can not only make better decisions but also avoid the problems that arise when their preferred management style causes them to make mistakes.

The best approach to regulate AI development is still ill defined. The usual approach of the government is to analyze and recommend based on good practices. With the rapid development of AI, this is a slow process. The need for an "adaptive organization" however requires the approach from a 'complex' state. Learning by experimentation to sense what is emerging in order to respond accordingly. Snowden & Boone (2007) also describe the Agile methodology as an approach that makes the switch between the complex and complicated domain. (ibid)

THE CYNEFIN FRAMEWORK FOR DECISION-MAKING

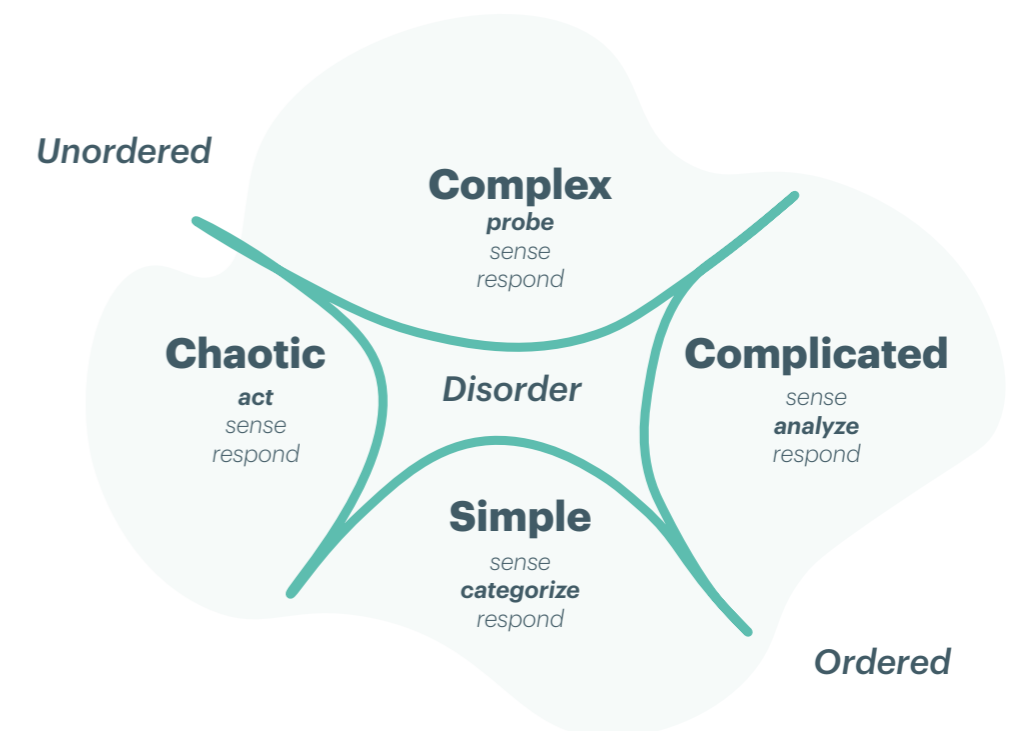


Figure 14: Visualization representing the leader's framework for decision making (Snowden & Boone, 2007)

Combining the Cynefin framework with the literature on Wicked Problems, we can conclude that these societal complex challenges should be tackled using the corresponding steps: Probe, Sense and Respond. The combination is visualized in Figure 15.

DECISION-MAKING FOR WICKED PROBLEMS

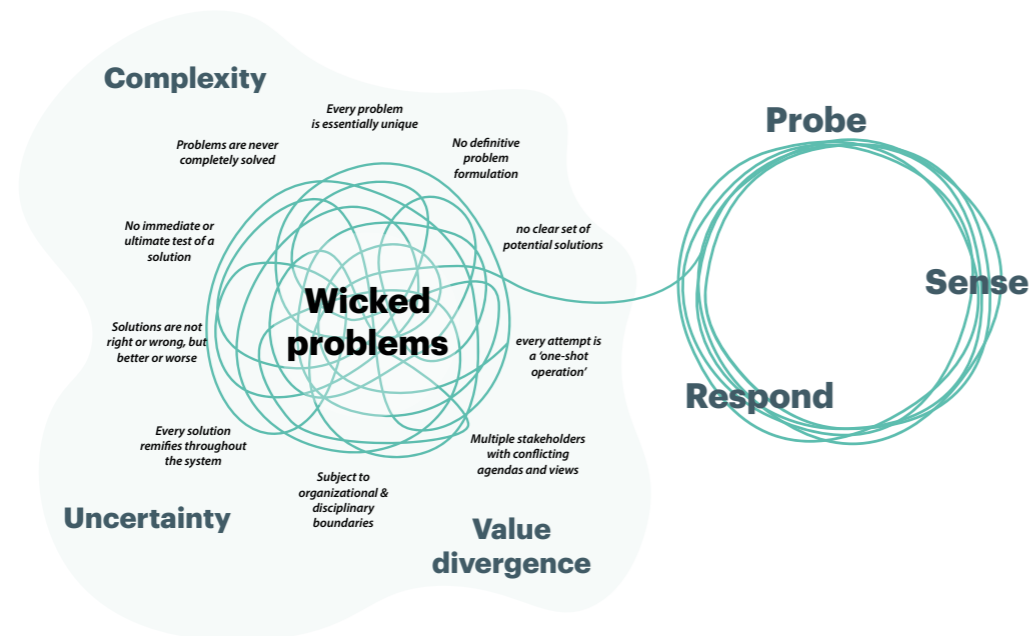


Figure 15: Combining Rittel & Webber's characteristics of a wicked problem with Snowden & Boone's framework

Most situations and decisions in organizations are complex because some major change introduces unpredictability and flux. In this domain, we can understand why things happen only in retrospect. Instructive patterns, however, can emerge if the leader conducts experiments that are safe to fail. That is why, instead of attempting to impose a course of action, leaders must patiently allow the path forward to reveal itself. They need to probe first, then sense, and then respond, representing an agile approach.

2.2.2 The role of the media

Our differences are magnified by social media and the government is under pressure to find solutions for pressing problems that are growing ever bigger. Public empowerment is on the rise. Not only is 'Participatory design' identified as a growing field in research publications, the expression of an individual's opinion can quickly gain attention, pressing urgency on the political agenda. Publications on social media can become viral in seconds, get picked up by the news, resulting in the pressure on the government to formulate a political statement in response. Big tech giants currently set the stage for ethical development. The recent Black Lives Matter movement triggered IBM, Amazon, and Microsoft to publicly announce they would not provide Facial Recognition technology to the police. Figure 16 highlights some recent news articles considering the negative consequences of using AI.

MAGNIFYING GLASS OF THE MEDIA ON AI

Figure 16: Snapshots of news articles presenting the magnifying glass

"Now policymakers only start 'walking', when the minister tells them to. Because the minister has promised something. That is now the reality." - Interviewee

Figure 17 describes the media as intermediate between the public service and society. It shows the current driver for policymaking, being the influence of the media.

"A vision of facial recognition will not come until the newspaper has written that the police are experimenting with it, which leads to concerned questions asked in the Second Chamber, as a follow-up of the publication in the newspaper. Then the minister promises: "You will receive a letter and we will clearly state what we are experimenting with and why." And then the public servants will have a few months to write a letter about it. Then they will try to get a kind of overview in no time and based on that overview, they will arrive at some kind of policy or vision. "Yes, then it will immediately be used as an argument for what he already wanted. That comes only after it has been in the newspaper. That is making reactive policy. The pressure comes from outside." - Interviewee

As all these parties exert pressure on policymaking, it calls for the need for stakeholder inclusion.

THE PROCESS OF REACTIVE POLICYMAKING

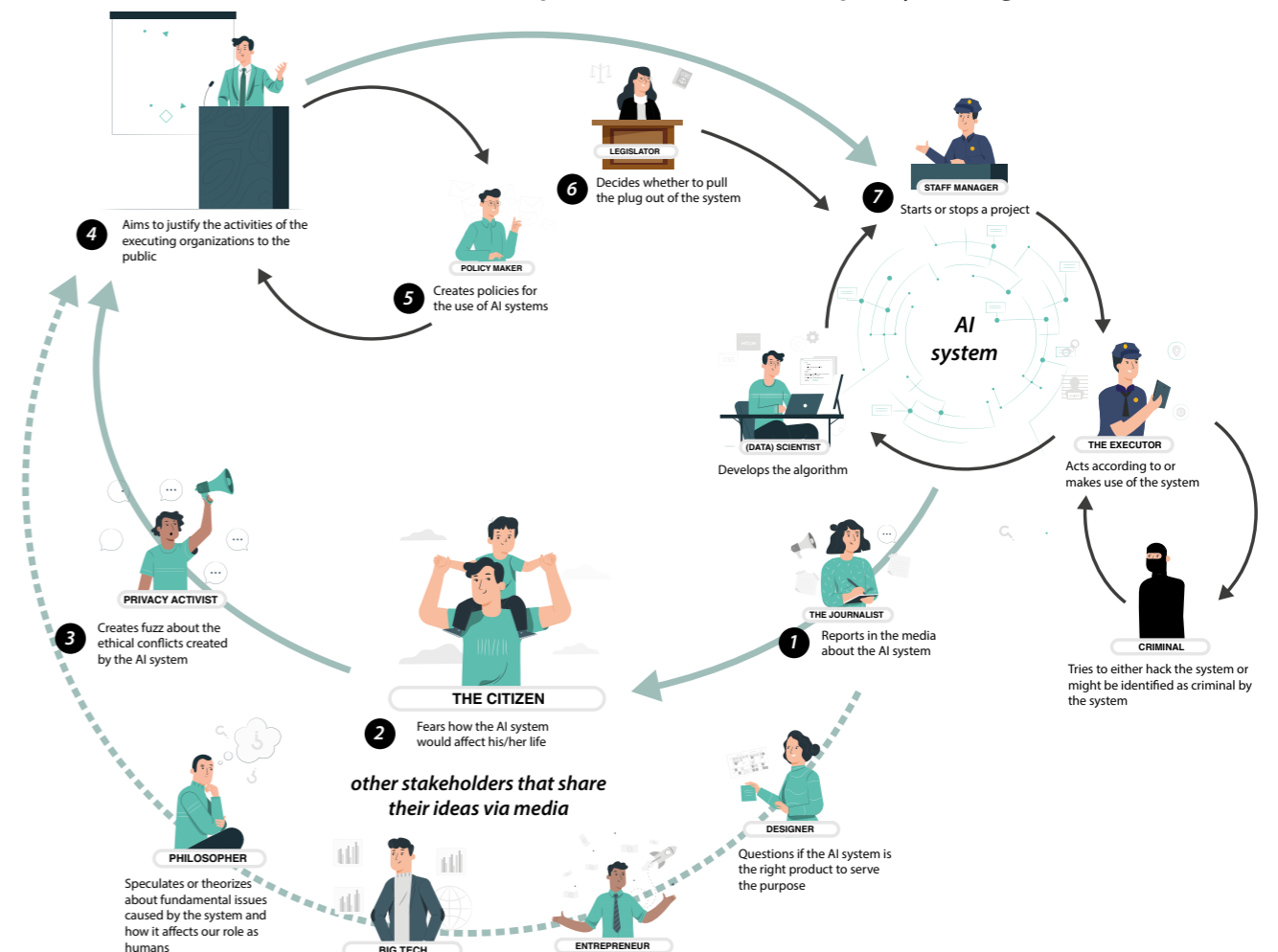


Figure 17: The process of reactive policymaking starting with publicity of AI projects

Achieving stakeholder inclusion with democratic renewal

Tallacchini (2009) argues that “good governance of research should take into account the need and desire of all stakeholders to be aware of the specific challenges and opportunities raised” with the terms ‘need’, ‘desire’ and ‘awareness’ gravitating in an unspecified way around the descriptive/ prescriptive concept of governance; and that inclusiveness should “allow all stakeholders to enrich the preliminary discussions” (Tallacchini, 2009)

Calls for democratic renewal are gaining more attention - on the street and in the media. Countries are already experimenting with new forms of decision-making, in which citizens are given direct responsibility for policy. The growing demand for inclusion by citizens becomes visible in the growing number of initiatives ranging from online polls to co-creation sessions with municipalities. Thus, new combinations of consulted expertises and perspectives in the execution of public services are growing phenomena.

One of those forms of democratic renewal is the “citizens’ council”. In October 2020, a motion by eight political parties was passed in the House of Representatives, which entrusted the cabinet with the task of examining the pros and cons of such a citizens’ debate in this case on climate policy. A citizens’ council is a way of crowdsourcing political decision-making, in other words: having it come from society. A group of drawn citizens makes policy recommendations based on extensive information and consultation. It is a form of “deliberative democracy”, democracy based on consultation and decision-making between citizens. The participants in a citizens’ council form a cross-section of society, including about as many women in it as men, the age structure and cultural diversity reflect that of society, there are many people in it with a practical education, some less with a university education, and they come from both town and city.

Other variations that have been widely expressed by the media are the rise of Break Out Teams, Multidisciplinary Intervention Teams, Bellingcat, and a new introduced method of the TU Delft considering a Participatory Value Evaluation. The ministry’s minister stated that the new Multidisciplinary Intervention Team (MIT) is expected to grow in the coming years into a team of approximately 400 additional specialists to strengthen the approach to organized undermining crime. (Rijksoverheid, 2020) With the research collective Bellingcat citizens voluntarily help solve police cases using online sources such as social media and geolocation software. It transcended the status of a group of volunteers to becoming an institute and a source of news for major international media. Another recent example is the launch of a new method by the TU Delft (2020) for evaluating policy options with a large pool of citizens. The Participatory Value Evaluation (PVE) tool provides citizens the ability to give advice about a government choice, in an easy-to-access way. In an online environment citizens get to see which choice the government has to make, they get an overview of the specific advantages and disadvantages (or effects) of the options which the government can choose from, and any restrictions that exist (e.g. a restricted budget or a compulsory aim). Thereafter, they are asked what they would advise the government to do and explain their choices, providing the government a clear picture of their preferences and what the public considered.

“The government needs to make the necessary changes, including a whole-hearted commitment to stakeholder inclusion at the heart of policy development.”
- Head (2008)

Identifying blind spots

Figure 18 exposes that asking 10 blind experts can certainly provide a lot of insight, but successful execution requires to approach the problem as a whole. Supported by one of the interviews with legal experts, this phenomena calls for curiosity in the context and inviting the biggest critics in the room. Hence, inviting multiple perspectives, postponing our judgment and feeling comfortable sharing our actions and thoughts. It also requires an open mindset in which you give each other feedback and consider an individual’s blind spots. Taking action informed by empathy and observation is a domain where only optimists may tread; optimism allows the existing condition to be interpreted as status quo rather than static fact. When asking an expert to advise, while another team takes care of execution, it lacks ownership and continuity. Hence, stewardship enables both thinking and doing in synergy.

NEED TO APPROACH THE PROBLEM AS A WHOLE

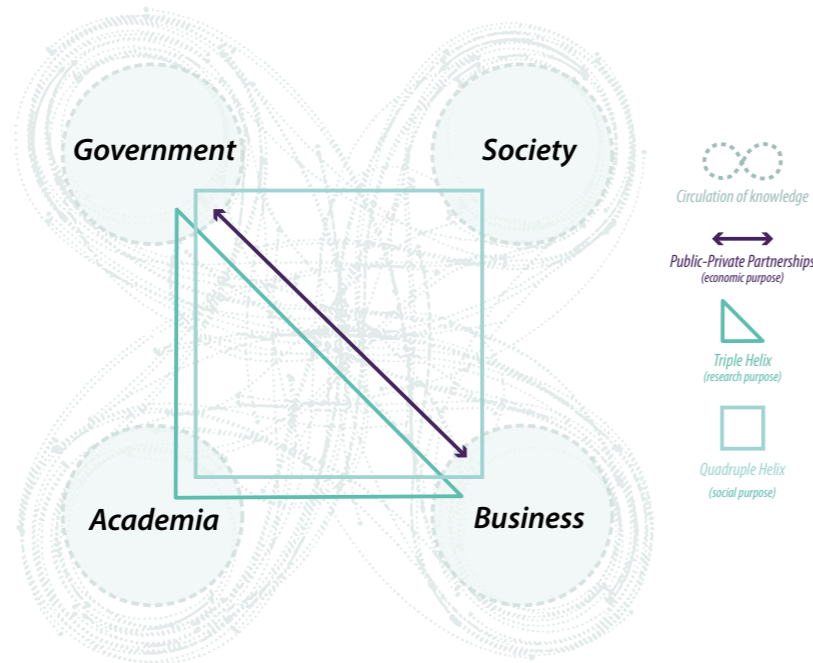


Figure 18: Experts viewing parts of the problem

The Quadruple Helix model

The Quadruple Helix model by Carayannis & Campbell (2009) incorporates the public via the concept of a ‘media-based democracy’, which emphasizes that when the political system (government) is developing innovation policy to develop the economy, it must adequately communicate its innovation policy with the public and civil society via the media to obtain public support for new strategies or policies. This model, shown in Figure 19, distinguishes academia, government, business and society as components of a knowledge economy. Hasche, Höglund & Linton (2020) defined the difference between the public-private partnerships (PPP), the Triple Helix and the Quadruple Helix. Where the PPP usually resembles an economic purpose, the focus on the Triple Helix is on its institutional format. The Quadruple Helix positions the social purpose as the main goal, resembling a collaboration with the public sector, the private sector, academia, and society. Moreover, the best way to establish the acceptance of policies is to create societal ownership, by giving these different perspectives the opportunity not only to be involved, but to participate in the co-creation of policies. Hence, Figure 20 proposes a desirability panel fueled with experiences by the representative roles of the four domains to help to identify the blind spots in policymaking.

THE QUADRUPLE HELIX: CIRCULATION OF KNOWLEDGE



COLLABORATION WITH THE QUADRUPLE DOMAINS



Figure 19: The Quadruple helix representing the collaborations in the four domains and its circulation of knowledge
Figure 20: Presenting the idea of a desirability panel covering the four domains

2.2.3 Policymaking process

The act of policymaking includes stating a goal, defining means to achieve the goal and, in most cases, offers a timeframe in which this goal will be achieved. When a societal issue comes to the attention and reaches the political agenda, it triggers conducting an analysis and setting priorities. After the priorities are set, the policymakers determine (in consultation with other parties) which format of intervention by the government is most suitable for the posed issue. Good substantiation is of great importance in this regard, as decision makers and citizens will ask for the reasoning behind the policy decision. The policy is therefore here defined as: *“the decision of the format of intervention, to achieve a stated goal.”* The level of intervention can range between stewardship (no intervention) upto creating new legislation, visualized in Figure 21. (adopted by the UK Policy Lab, 2014). A middle way is for example creating an awareness campaign.

OPTIONS OF POLICY INTERVENTION



Figure 21: The options of policy interventions in relation to the level of involvement and scale

A policy regulation can indicate when an exemption or permit will be granted or not. For example, an interpretation of regulations can also be given through a policy rule. This is not binding for a judge. With the policy regulation instrument, the government explicitly chooses a position above the parties involved, which means that there is a vertical form of management. Policy rules are further unilateral, because no quid pro quo is required. In principle, they are generic in nature, because the government aims to treat people or social actors in the same situation equally. Policy rules can be seen as direct regulation (there is no question of financial, economic or social regulation). Policy rules are moreover directive in nature, because they are pre-eminently aimed at realizing behavioral changes in the shorter term.

The quest for future proof regulation

The act of policymaking and execution is often defined as a linear process, visualized in Figure 22. The linearity has the risk of withholding the response to changes, as passing the line through vertical stakeholders slows the process.

POLICY CREATION TO EXECUTION: A LINEAR PROCESS

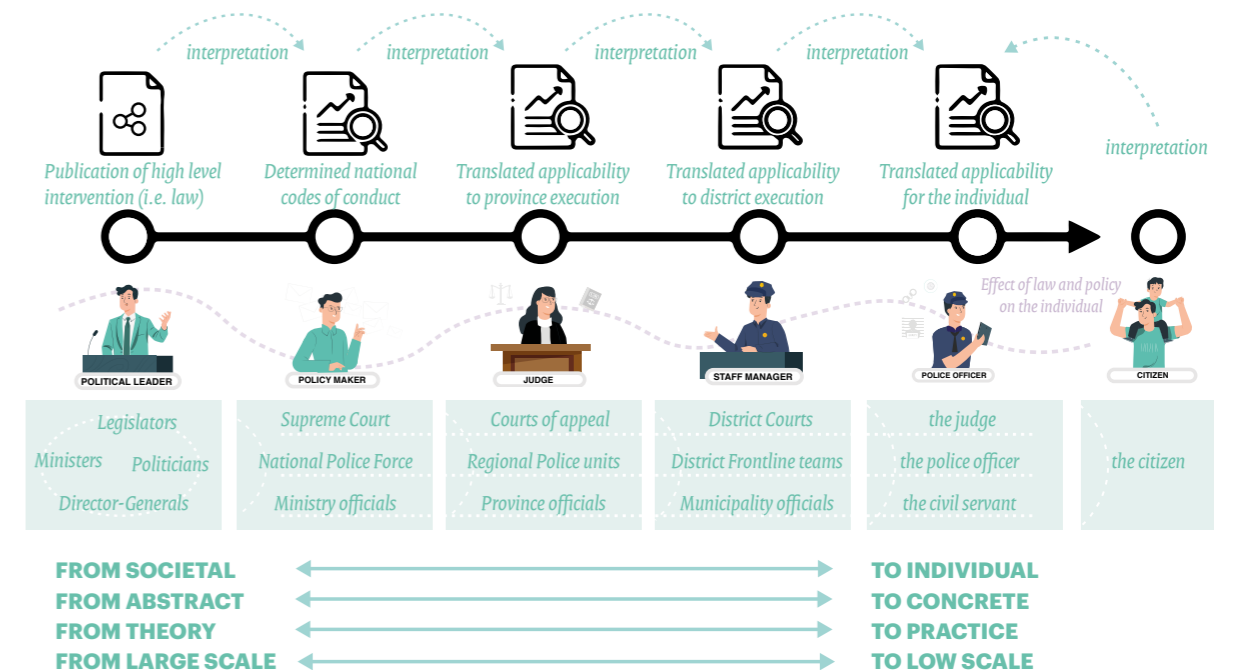


Figure 22: Presenting the linear process of policy execution

Policy development occurs across a range of contexts, from the settled routines of ‘business-as-usual’ through to ‘crisis management’. Under conditions of crisis, with strong pressure for immediate action, the need for conspicuous action may tend to overwhelm the opportunity for new thinking. Rather than a positive and thoughtful ‘paradigm shift’, there is a significant likelihood of reinforcing past practices, e.g. ‘group-think’ about tactical responses and use of top-down styles of emergency coordination (Boin, Stern & Sundelius, 2005).

New strategic thinking needs to be championed within the public sector. This requires organisational learning and cultural change. As the digital transformation is going rapidly, amendments to legislation are necessary to continue to safeguard public interests. “Nowadays there is a lot of attention for future-proof legislation, but that is no panacea: the legislator will have to remain alert to disruption and respond creatively.”(Mul & Werkhorst, 2020) “We must let go of the idea that laws have been established forever and must continue to look at this adaptively.” (ibid)

The level of responsivity to embrace new technology is hereby defined by the ability to:

- absorb shocks (game-changers), which asks for reflecting on the whole and identifying patterns and game-changers proactively
- flexible adjustment (adaptation), which asks for agile principles
- actively shape the future (proactive), which asks for a vision and a process of continuous learning by doing

One of the theories on this responsivity is the Cycle of adaptive change (Gunderson & Holling, 2002), represented in Figure 23. The cycle describes adaptive change as the continuous loop of the four phases: Conservation, Release, Reorganize, and Growth. The phase Release links to radical innovation, whereas the Growth phase links to incremental innovation. Mul & Werkhorst (2020) state that the core of this theory is that complex adaptive systems can walk unnoticed into the rigidity trap. As these systems tend to exploit ‘what works’ it leads to the conservation phase, where a system or company is tied to only do what it always did. At some point a ‘game changer’ presents itself, making small acts of reorganisation not effective enough. Signs become visible that the structures of established order will collapse (Release). A time of radical innovation begins, which creates room for a turbulent period of learning and experimentation. To secure the newly built capital and promising solutions that derived from this period of experimentation, a new solid foundation is needed, requiring a reorganization of the system (Reorganize). When a succesful solid new foundation is built, a path of incremental innovation (Growth) will lead back into the ‘business as usual’ (Conservation).

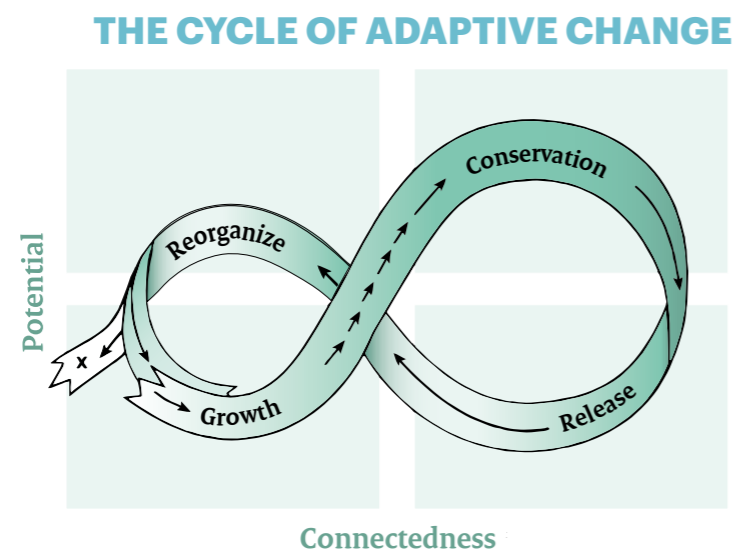


Figure 23: The Cycle of adaptive change (Mul & Werkhorst, 2020)

Primary types of policy disruptions by business

“Policy disruptions can result from conscious choices by entrepreneurs to exploit legal loopholes or to challenge regulatory protections for incumbents. But it can just as easily result from gaps in a regulatory regime or fundamentally new business models that solve problems legal regimes have been designed to address” (Biber et al. 2017) This quote emphasizes the need for explicit regulation, or else the amount of AI applications that exploit legal loopholes could increase, and so does the percentage of unintended negative consequences.

Biber et al. (2017) distinguish four primary types of policy disruption: End-runs, Exemptions, Gaps, and Solutions, each of which requires a different regulatory response.

- End-runs: Policy disruption caused by conscious choices by entrepreneurs to exploit ambiguous laws.
- Exemptions: Policy disruption caused by entrepreneurs expressing legal loopholes .
- Gaps: Policy disruption caused by business innovations to which the existing regulatory regime simply does not apply
- Solutions: Policy disruption caused by business innovations that could solve problems of which the regulatory systems were initially designed to address, but cannot be used as the model does not match the requirements or assumptions of the current regulatory system.

The public’s acceptability of policies

The political viability of the proposed regulatory responses mainly depends on whether it falls within the range of acceptability. Therefore the Overton Window was developed, that describes the extent to which we find policies acceptable (Szalek, 2013). The Overton Window regarding the Covid-19 pandemic is visualized in Figure 24. When a ‘game changer’ enters the field (e.g. Covid-19 virus), it appears that the window extends to the way that the unthinkable ideas actually become acceptable. So, as the government responds to the Covid-19, it’s preferred styles of action are changing. Preferable futures people may consider today, could be more divergent and more radical than before. The ‘Intelligent Lockdown’ process is an example of an agile approach to policy formulation. The key insight is that people want to know what they are up for. During a crisis, you want to experience some sort of control. Similar to a train delay, it’s more acceptable if you’re updated on the delay in cause and time. Therefore, each month the rules are revised, based upon monitoring the crowd and how people follow the rules. At the end of the month, the press conference is held to communicate the current behaviour and the consequences regarding softening or sharpening the rules.

THE OVERTON WINDOW REGARDING COVID-19

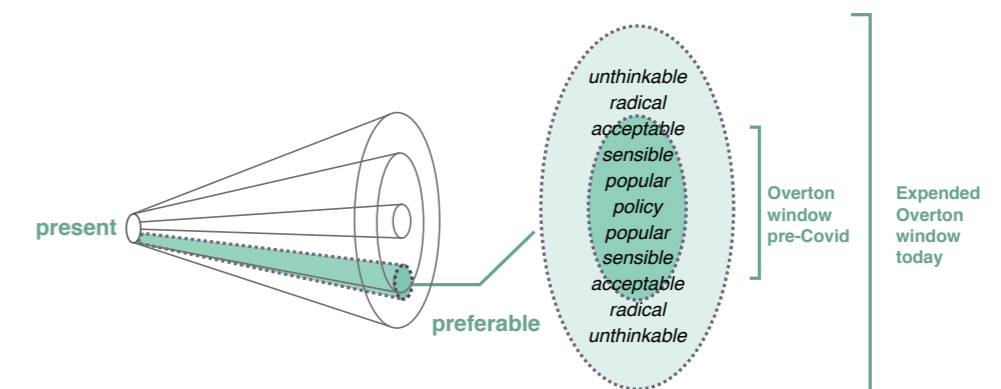


Figure 24: The Overton Window regarding the Covid-19 pandemic (adopted from UK Policy Lab)

2.2.4 Agile policymaking

Based upon the theory of 'Opgave Gericht werken' and the interviews with multiple policymakers, it can be stated that '95% of current policymaking is focused on the short-term, acting upon current issues that rise by the society's demand. However, there is a need for an adaptive culture at the government, especially considering the policymaking. The incremental belief is that you need an adaptive approach to solve wicked problems. Agile principles can be used as a valuable inspiration to define a certain adaptive 'character' in the process of policymaking. (Van Delden et al., 2020)

Translating the Agile Manifesto to policymaking

The four fundamental principles of the Agile Manifesto translated to the act of policymaking result in the first 'agile policymaking' principles:

1. Focus on customers over contract negotiations

Policies should focus on the end users, which is a dual-customer. Satisfy the minister and satisfy the citizen. To do this, policymakers need to understand user behavior in detail and design policies to address their needs. Policymakers should not delegate this step to implementing agencies, but learn to empathize with the user themselves.

- Principle: Specify the starting point of the intended person. Investigate the latent need on which the policy should be formed. Ask 5x why to get to the essence and the answer contains a person.

2. Prioritize working software before documentation

A reference to the art of maximizing the work not done. "A designer knows he has achieved perfection not when there is nothing left to add, but when there is, nothing left to take away" (Antoine de Saint-Exupéry). Focus on optimizing the system instead of the tasks. Reflect on what is possible to automate and what is the minimum viable way of working. Invest in the state of the art co-creation software with the aim to easily share, adjust, iterate upon policy documents. Bureaucratic processes of excessive documentation can limit the ability of teams to adapt to change. An important consideration in this matter, however, is the ability to explain to the minister how certain regulations and policies are created and decided upon. The current approach consists of elaborate textual documentation to cover this ability to verify and recall the decisions made. So there is an option to rethink the possibility to recall decisions with minimal effort.

- Principle: Document only minimum accountability requirements. View alternatives for long text documents. For example, use programs and systems with automatic backlog or co-create with visualisations that entail rich layers of information

3. Encourage individual interactions, not just process

Rapid and regular informal communication is more important than formal bureaucratic communication. Verbal communication that is early and frequent will help address barriers and challenges as soon as possible without the need for lengthy memos and documentation. When formulating policies, this asks for structural alignments with the relevant horizontal, vertical, cross-departmental and external stakeholders. Create multidisciplinary teams, consisting of both policy makers and implementers, that act in a network bridging both horizontal and vertical teams.

- Principle: Deliver intermediate results in short cycles, above final result. Organize structural check-in moments (per month) with a diverse-oriented core team (i.e. Multidisciplinary Intervention Teams)
- Principle: Face-to-face coordination in vertical and horizontal teams. Organize cross-dimensional meetings where policy such as implementation as critical partners offer each other's perspectives, so that over time you learn to speak the same language.

4. Plan for change instead of following a plan

Expect, welcome, and adapt to change. "Current policies are formulated with the mindset 'is it safe?' instead of the driver statement 'is it safe to fail?'" (interviewee). Instead of preventing laws that do not leave room for experimentation, we can approach this process by making explicit what we certainly don't want by referring to specific scenarios. Design plans and policies expect to prove your assumptions false at some point. The earlier you recognize change in the policy users' needs, the earlier you can respond to change. Create rapid feedback mechanisms to better understand the change along the way. Do not wait for end-of-program evaluations to seek feedback on policy performance.

- Principle: Start with trial period. Provide a safe environment to fail.
- Principle: Describe law as goal and test contextual micro-rules in parallel. Define the goal as a guide, but write policy for exploration where the effect can be reflected within a defined time frame
- Principle: Reflect on both process and result, short and long term. Make use of double loop learning: both the process and the result offer underlying insights.
- Principle: Innovation as a task, adaptivity as a goal. Policy should not be continuously adjusted, as long as the goal remains in sight and is critically reflected upon.

Methods for agile policymaking

Regarding the formulation of policies, the tough challenge is that these arise within a political environment. These conversations aren't about being right or wrong, but about the choice to agree or disagree. Whereas a policy is aligned with the opinion of the mass, a product is tailored for the individual. Correspondingly, whereas you choose to buy or lease a car based upon your own values and preferences, a policy formulation on when and how to use your car covers different angles of interpretability. A policy option to reduce the eco footprint could be to lower the speed limit, but others might disagree as they value time saving over sustainability. Such a complex ecosystem of human behavior is hard to grasp in a physical product, so the remaining question is: How can we?

With the aim to limit the timespan of the policy formulation as technology gets ahead of us, the focus should be on the ways to gain strong evidence and which require a short time to set up and conduct experiments. Figure 25 shows various methods to experiment, plotted on the axis of gaining strong evidence and the timespan to conduct experiments.

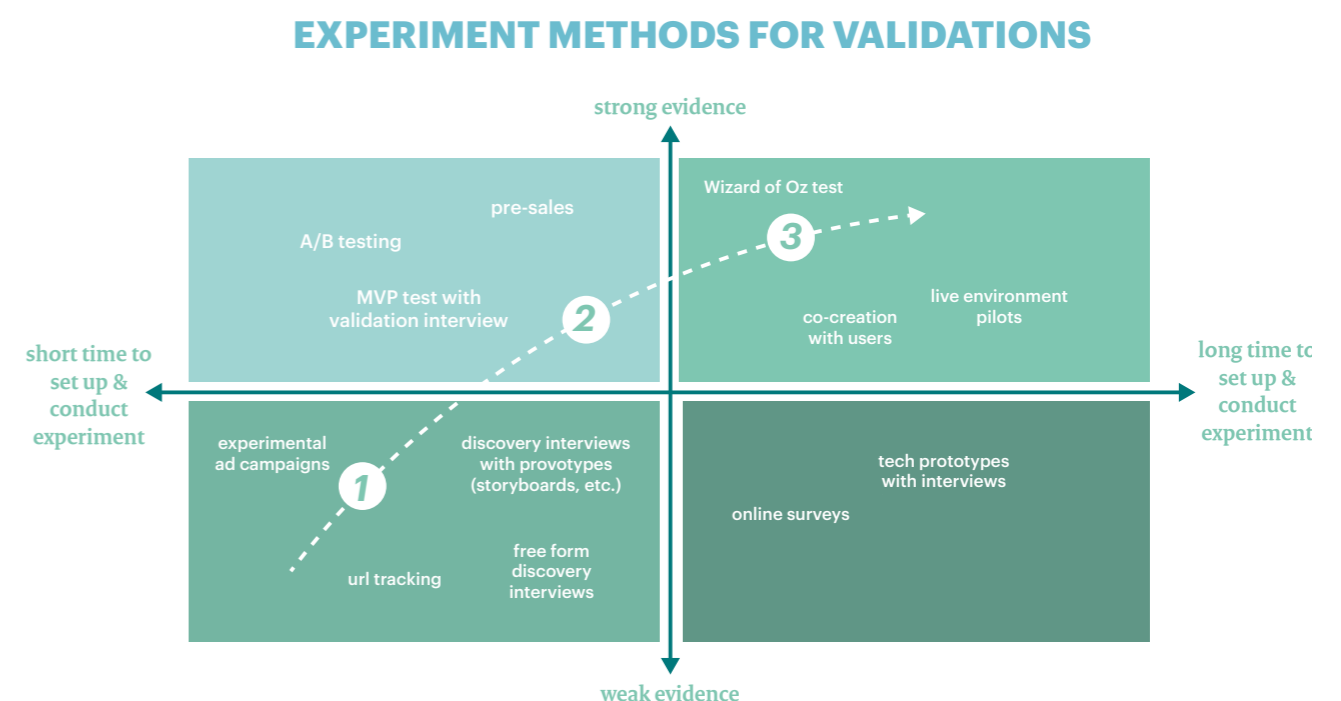


Figure 25: experiment methods plotted on required timeframe and quality of evidence - building upon the visualization adopted from Kolk (2019)

To clarify the areas 1 to 3, a method representing the area has been translated to the act of policymaking. Figure 26 describes the three phases, featuring: the Provotype, the Prototype, and the Pilot. A 'provotype' can be seen as a design intervention: "a form of inquiry that is particularly relevant for investigating phenomena that are not very coherent, almost unthinkable, and under-specified because they are still in the process of being conceptually and physically articulated." (Halse & Boffi, 2016) It intervenes expectations and models of thinking by investigating possible future scenarios, new concepts or problematic side effects of existing services. These are great in provoking discussion because they visualize and concretize ideas much further than spoken words or written descriptions.

FROM ABSTRACT TO TANGIBLE IDEAS

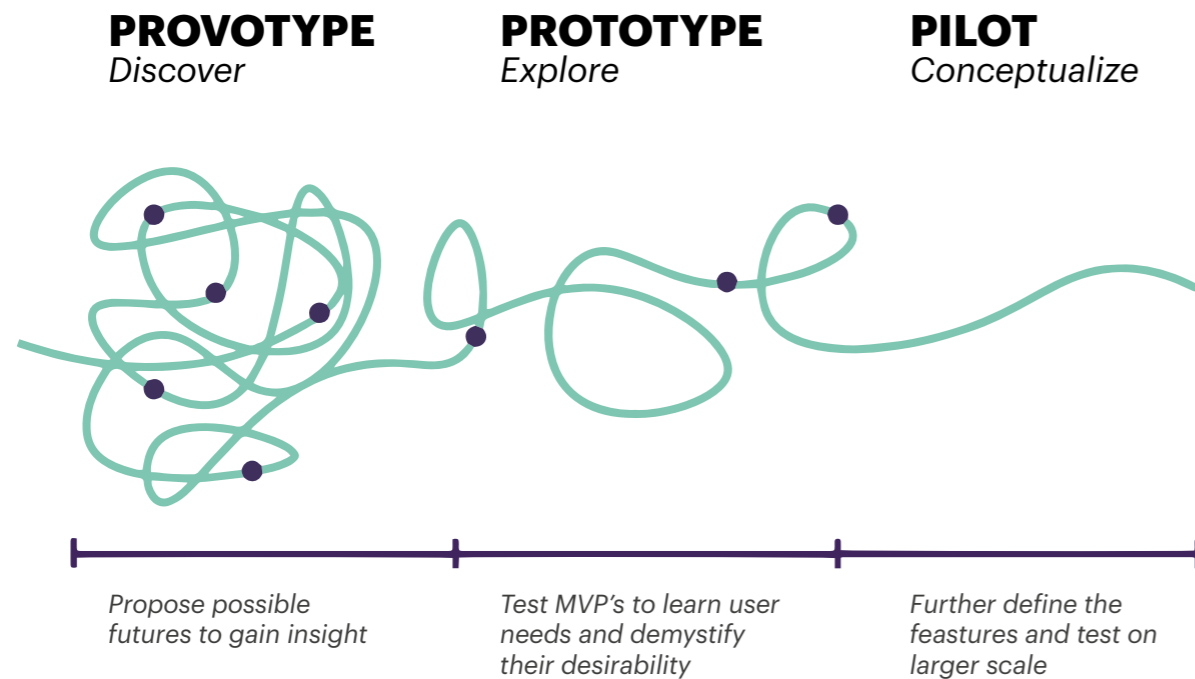


Figure 26: differences in the purpose and fuzziness of a provotype, prototype and pilot - builded upon the visualization by Kolk (2019)

Depending on the phase whether the aim is to discover, to explore or to conceptualize, the act of policy making could follow this explorative structure. What if the government would create Policy Provotypes and the public could respond to these initiatives? This act resembles gathering feedback fast, open and in full transparency. The first gained insights would steer how to frame the policy, which could lead to the creation of the prototype: the 'Minimum Viable Policy'. This would be a first attempt at a stated regulation, where an organization should be involved that would volunteer to test the new policy, and measure its effect. This will generate a proof-of-concept. After that, enough insights are collected to invest in a long-term Policy Pilot. This pilot could be a regulatory sandbox, where a regulator allows developers to conduct live experiments in a controlled environment under a regulator's supervision to observe and validate the reaction on a proposed policy option. This would require a larger amount of investment in time, money and resources to be able to execute the test. This could be the last step before implementation, as you took the citizens along in the process and therefore reduce the fuzz around it when the policy is made definite.

The policy choices can be prototyped and tested with key users and the identified potential affected stakeholders. You can test multiple policy ideas at once to understand the strengths and weaknesses of each choice. Testing does not need to be excessively scientific and rigorous, the main goal is to test your assumptions in the real world. Get the policy in front of the users who matter most and understand its impact on their operations and behaviors.

Leeuw (2019) proposes the use of micro-directives with an example of traffic regulation: A law now contains norms or standards (i.e. 'showing reasonable driving behavior'), while a rule contains a concrete SMART* formulation (i.e. 'maximum speed 60km/h on this road between 6am-6pm'). The following examples help to understand the concepts of the Policy Provotype and the Policy Prototype (i.e. the Minimum Viable Policy).

A Policy Provotype could be a provoking story to discover the range of acceptability. For example: Introducing tailored laws for traffic rules. Feedback on this policy option can be collected via testing the provotype. To illustrate the concept, see Figure 27.

POLICY PROVOTYPE EXAMPLE

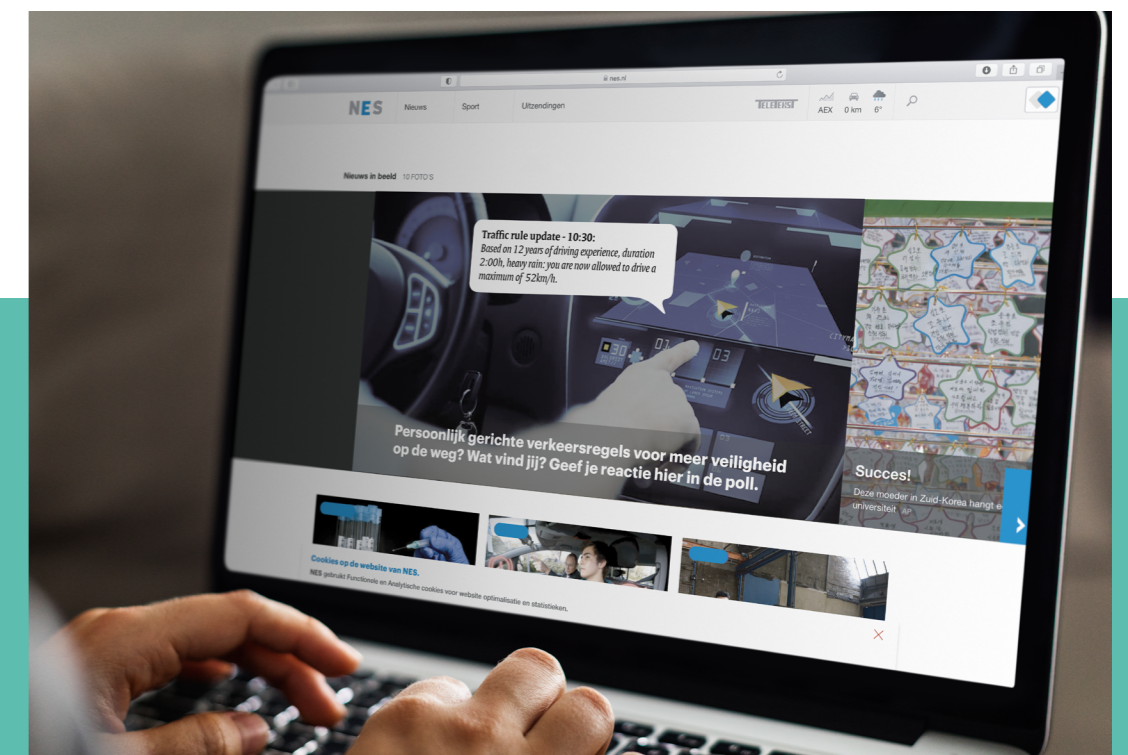


Figure 27: example of a Policy Provotype test-setup

Imagine a news source that has a regularly updated poll with future scenarios. These scenarios depict a hypothetical new policy and ask the reader to post their reaction and/or describe an alternative policy option. The format of the poll could be the following text with a concept visualisation:

"Traffic micro-directives to improve road safety by reducing road accidents. An algorithm created a catalog of traffic rules using the following context factors: driver's license years, duration of travel, and the weather condition. A micro-directive pops-up on the driver's dashboard: "based on 12 years of driving experience, duration: 2:00h, heavy rain: you are now allowed to drive a maximum of 52km/h in this area. Would you consider it fair if another driver is allowed to drive 20km/h faster? What do you think? Leave your comment below"

A Minimum Viable Policy could be defined as a policy draft version with several policy intervention choices to consider. The concept includes a test set-up to measure the effect of potential policy choices. Instead of asking what feeling the potential scenario would evoke, the prototype is representing a real-time case in which the responses are observed. A Minimum Viable Policy test set-up could range from a fake camera and poster in a supermarket stating “We use emotion recognition to provide you a tailored discount on your groceries” to a headline in an unexisting newspaper, to trigger the public’s reaction. To give an example of a Minimum Viable Policy, a concept is illustrated in Figure 28.

MINIMUM VIABLE POLICY EXAMPLE



Figure 28: example of a Minimum Viable Policy test-setup

Imagine a potential policy option would be to combat shop theft using pattern recognition in stores. Instead of developing the software to test its feasibility, a poster can be created in minutes to test its acceptance (as well as desirability) first. After a week, measure the effect of the traction on social media.

2.2.5. Insights & Conclusion

THE BUILDING BLOCKS



Experience-driven:

Only inviting experts in the room won't help. The ones who experience the effect on AI systems in daily life can bring valuable insights. Include various fields of knowledge to enrich problem-solving. Wicked problems can only be addressed by identifying the 'elephant' as a whole.



Empathy:

Understanding policy issues starts with empathizing with the affected stakeholders. Build bridges by validating your assumptions early in the process. Observe to reveal latent needs, ranging from insights on why current ethical tools are not attractive to use, to understand what concerns various stakeholders have and what power they have when they don't feel heard.



Inclusive:

Create policies with a representative group of society to bring transparency to the public and secure the acceptability of public service that includes AI software. Combine a citizen panel, with public-private partnerships, and the knowledge of academic research, representing the Quadruple Helix. Successful policies start with identifying the blind spots of the actors in the act of policymaking.



Iterative:

Wicked problems ask for a probe, sense, respond structure. Similar as the acknowledged desire of the ministry to shift to a task-oriented approach (known as 'opgavegericht werken'), approaching regulation with sprints will help the government to become more adaptive and resilient.



Future oriented:

95% of current policymaking is focused on the short-term, acting upon current issues that rise by the demanding society. Current approach is resembled by reactive policymaking. Hence, policymakers need a momentum to reflect on the long-term goals instead of the issues of the day to achieve organizational learning.

2.3 Current state of AI experimentation

In this section, the spectrum of governmental experimentation is clarified, the relevant stakeholders are examined, and the collaboration within the SBIR use cases are explored.

2.3.1 Spectrum of experimentation

Research question #4 considers the necessary building blocks to enable organizational learning. The driver behind this question resembled the ministry's desire to 'learn by doing.' Hence, learning about AI by conducting AI experiments. Governmental experiments are deployed through various channels, depending on (1) the purpose of the experiment and the phase, and (2) the level of involvement. An overview is shown in Figure 29.

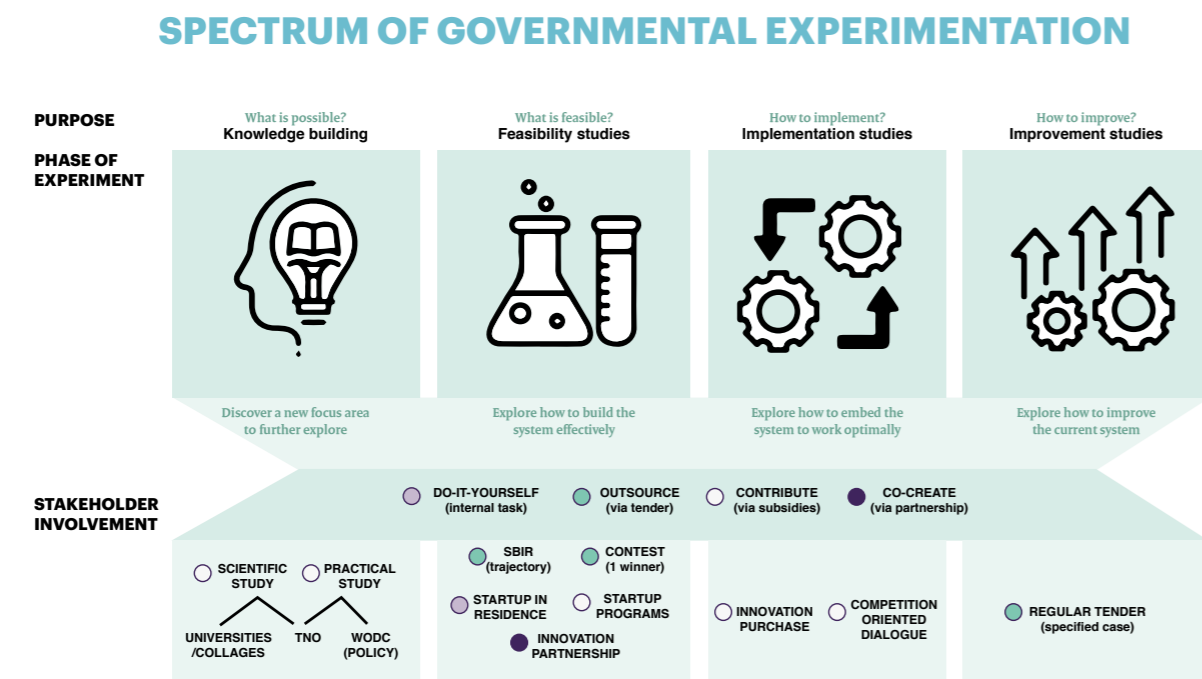


Figure 29: The spectrum of AI experimentation fulfilling the purpose of knowledge building, experimentation, implementation and improvement

1. The purpose and phase

The aim of starting a research or experiment is related to four phases:

- knowledge building (what is possible)
- feasibility studies (what is feasible)
- implementation phase (how to implement), and
- the improvement phase (how to improve)

Knowledge building has an explorative character where studies are conducted usually only to retrieve new insights. When the research has a scientific purpose, it is usually done in collaboration with universities and colleges, whereas a practical purpose study is usually done in collaboration with the WODC. This is a knowledge institute in the field of justice and security, which independently conducts or outsources scientific research for policy and implementation purposes. The ministry also closely collaborates with TNO that covers both scientific and practical purpose studies.

When the possibilities have been explored, the next step is to decide on the solution space. This is where a feasibility study shows which innovative ideas have the highest potential to be realized. This phase initiates to start experiments. The starting point here is what innovation is a feasible solution to a proposed challenge. Such feasibility studies follow a proof-of-concept, a real-life experiment (also known as a 'proeftuin') and/or a low scale pilot. In the implementation phase, an innovation purchase can be made or a competition oriented dialogue can be arranged. In the improvement phase, a regular tender is sent for a specified challenge.

2. The level of involvement

The ministry can either choose (1) to outsource, (2) to contribute or (3) to start an experiment themselves. When decided to outsource the project, a tender is created. According to European guidelines, governments are obliged to publicly tender for upcoming deliveries, services or work that exceeds a threshold amount. This means that they must announce that they intend to purchase goods or services, after which companies can tender for the contract. Because there are more interested parties, the chance of an economically advantageous supplier is greater than when one supplier is chosen directly. This also avoids nepotism (in Dutch known as 'vriendjespolitiek'). The tender can have the format of a contest leading to 1 winner, or the Small Business Innovation Research (SBIR) format (see Figure 30). The SBIR process is a phased innovation competition, in which the companies with the best offers go through each phase. SBIR is the government's purchasing of research services in order to create a market in which the government can later make their own purchases. A contracting authority identifies a specific challenge and makes a budget available. The entrepreneurs with the most promising feasibility studies (phase 1) are instructed to further develop their product (phase 2). After that, the entrepreneur can bring the innovation to the market himself (phase 3), the government can be the first customer ('launching customer') to purchase the innovation.

SBIR PROCESS PHASES

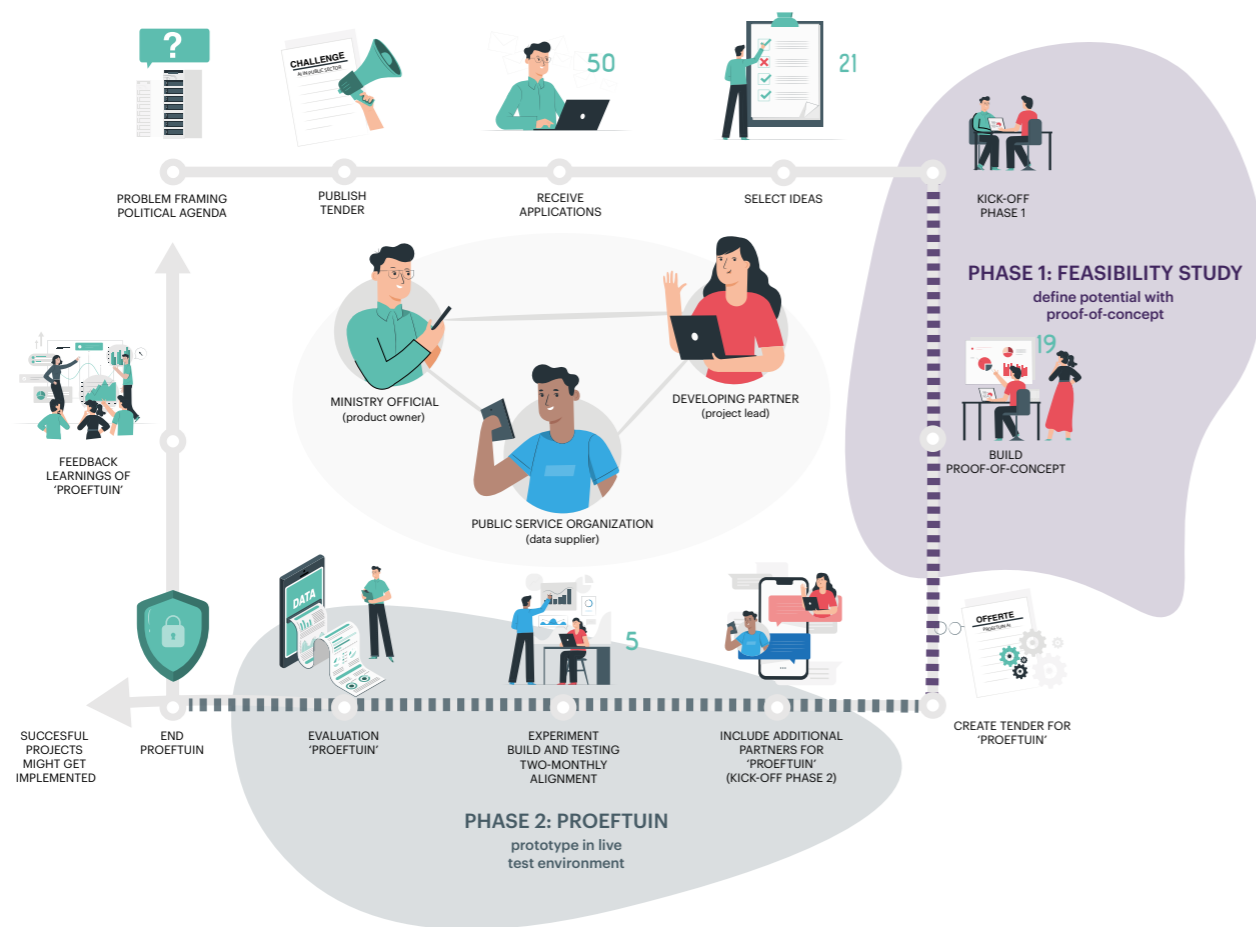


Figure 30: The SBIR process visualized

The aim of these experiments is the ability to improve the learning capacity. This emphasizes the need to reflect on these experiments.

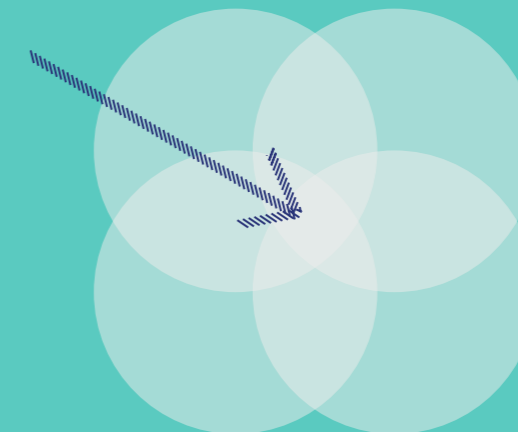
Contributing to an experiment is mostly done by means of providing grants, which are often provided to startup programs. Innovation partnerships require a co-creation level of involvement.

When decided to start an experiment by themselves, the challenge is included in the Startup in Residence (SiR) program. As start-ups take a fresh, open look at the (social) challenges the ministry faces, the program teaches them to look at issues in a different way. This cooperation leads to concrete, applicable solutions. It also provides room for startups to propose innovative solutions to the ministry without being asked, which is stated to be experienced as very inspiring and educational.

What we learn from these processes is that they are designed for a focus on assessing feasibility, whereas the desirability aspect is not publicly evaluated. When the government examines the desirability of an AI public service after the feasibility study is done, it risks a lack of understanding, support and a feeling of ownership. This might lead to protests, and the need to pull the plug out of the project. When taking desirability as a start, it will first create alignment and support, which smoothens the feasibility of a project.

“The sweet spot of business innovation is the intersection between desirability, feasibility and viability.” - Calabretta (2016)

“The sweet spot of governmental innovation is the intersection between desirability, feasibility, viability, and allowability.” - author (2020)



2.3.2 The lenses of governmental innovation

Calabretta (2016) defines the sweet spot of innovation as the intersection between desirability, feasibility and viability (see Figure 31). This theory could be best defined as a perspective: “a lens through which one can observe markets, customers, needs, competitors, and customer segments differently, and by doing so, make innovation far more predictable and profitable” (Ulwick, 2017).

LOCATING THE SWEET SPOT OF GOVERNMENTAL INNOVATION

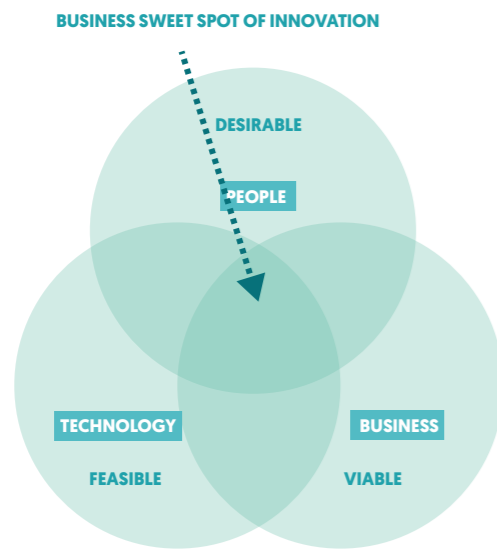


Figure 31: Sweetspot of business innovation

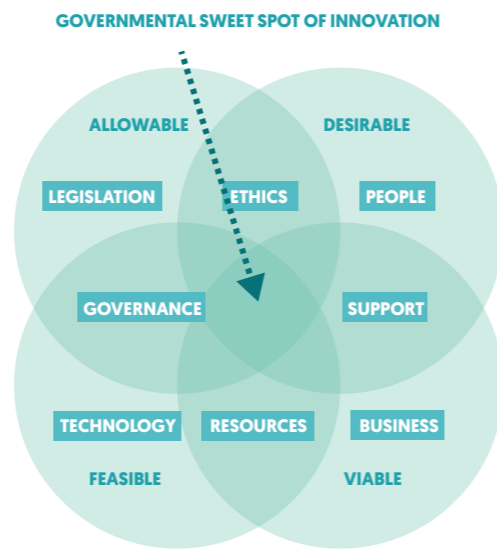


Figure 32: Sweetspot of governmental innovation

“Design thinking is a human-centered approach to innovation that draws from the designer’s toolkit to integrate the needs of people, the possibilities of technology, and the requirements for business success.” (Tim Brown, Executive chair of IDEO) Hence, the Design Thinking approach builds upon the theory on the sweet spot of business innovation. The lenses represent a balance between business (viable), people (desirable), technology (feasible), in which the overlap resembles the sweet spot. (Calabretta, 2016).

However, business innovation differs from governmental innovation as the lens of allowability (considering the norms, laws and regulations) is seen as an external influence or a ‘given’ for businesses, while the government has the power to take it into consideration. Hence, this lens is added building upon the theory for the sweet spot of governmental innovation, represented in Figure 32.

Empirical research described in the previous sections, revealed the current way of innovating is raised by the technology-push and the pressure on the government in how to control the rapid tech development. If there is one thing a policymaker fears, it is persistent uncertainty. The public asks for transparency by the government, while the image of the government is experienced to be threatened when the government would openly mention that they don’t know the answer.

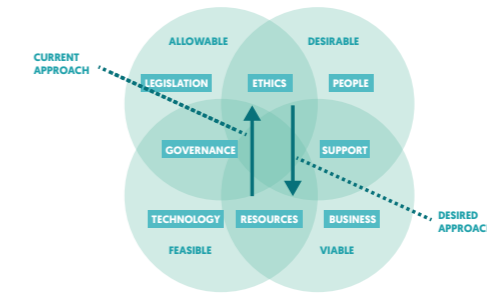


Figure 33: resembling the shift from the chronological order of questioning ‘what is possible’, to ‘what is desired’

THE ORDER AND INCLUSIVITY OF GOVERNMENTAL INNOVATION

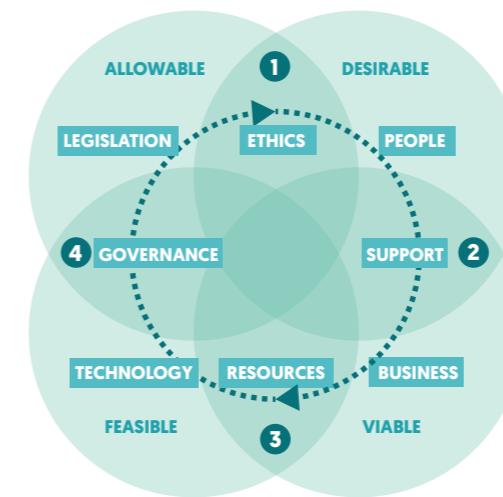


Figure 34: The order for using the lenses of governmental innovation

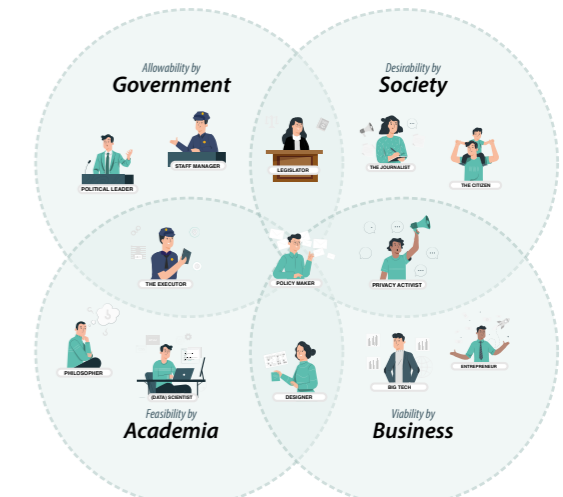


Figure 35: An inclusive panel representing the Quadruple helix domains

What is currently missing, is the order in which the lenses are used. Starting from the desirability lens, an AI application has its fundamental creation derived from the values the application should pursue. This approach is opposite to the current approach, shown in Figure 33.

When there is value-alignment by the relevant stakeholders (people), it can gain support from both internal and external parties to further pursue the idea. When the risks of the application are identified in order to make the application resilient (i.e. sustaining) against threats (e.g. intended misuse or environmental business factors) and the organizational infrastructure is build to lend itself for the application, the data resources can be accessed (due to the early support for including the relevant people) and quality can be enabled. As the threats (e.g. misuse) are limited by prior value-alignment and consciousness of the potential risks, the less intended data pollution is expected to occur. The algorithm can be trained while the governance is clearly indicated (responsibility, transparency etc). When the experiment then conflicts with current legislation, the need to adopt/change either the system or the regulation can be proposed. These lenses also build upon the previously mentioned Desirability panel, representing the need to include the representatives of the four domains to empathize with these perspectives through their lenses. The order and resemblance with the domains are shown in Figure 34 and 35.

2.3.3 Identified challenges in AI experiments

For the relation to the AI experiments, three “proeftuinen” that are part of the SBIR process have been explored, to understand the challenges faced in this process.

Case 1: deployment of resources
Spraklab: Audio speech recognition and filtering of relevant sentences

Case 2: inspection and enforcement
Oddity.ai: Violence detection in public spaces using image recognition with camera surveillance

Case 3: detection
Ziuz: Detecting incriminating material and filtering of video footage

These use-cases have been examined by interviewing the involved startup founders using a generative tool to explore their positive and negative experiences regarding the collaboration with the government throughout the process of developing the AI system. Figure 36 depicts the created generative tool.

GENERATIVE TOOL TO REVEAL EXPERIENCED CHALLENGES

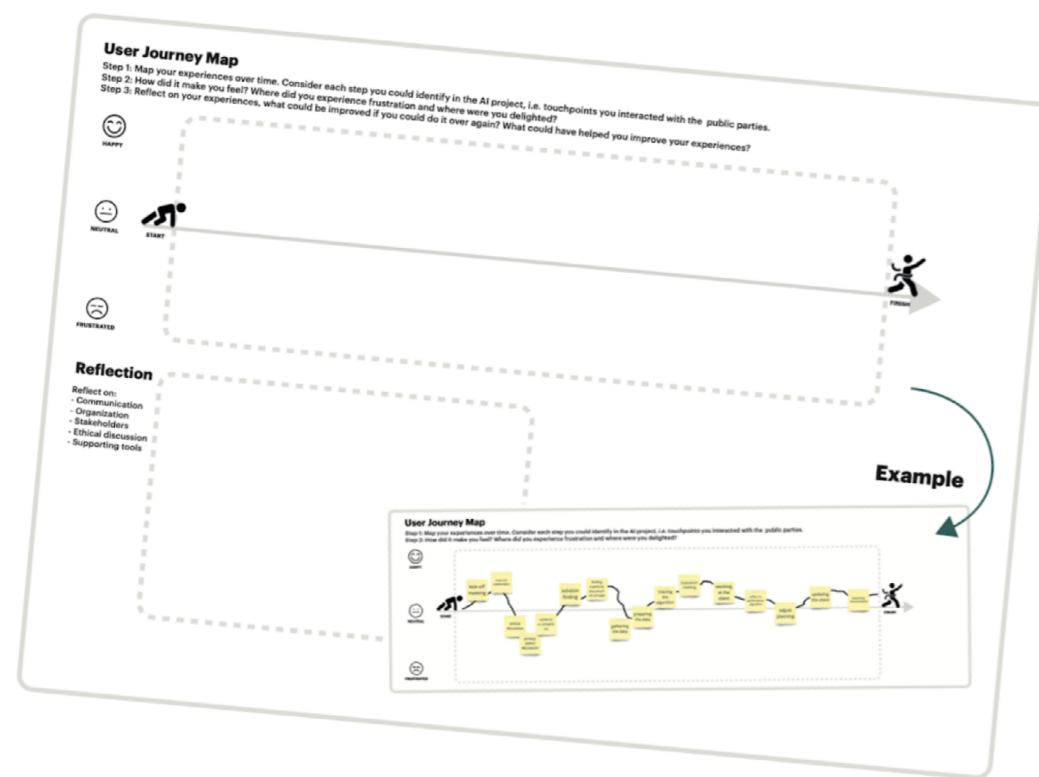


Figure 36: The created generative tool to elicit the experience of AI startups

The main gained insights are listed:

A) Desire for guidance

- “I need some guidance or guidelines on how to deal with this ethical discussion. Also because it helps me to show external parties that we are thinking about it” (interviewee)
- “Current tools are immature, current practices are experimenting which tool does and doesn’t work. There are many initiatives of ethical tools, it’s not clear which one to choose.” (interviewee)
- “If you would have asked me one year ago, I would not have interest in an ethical tool. Now I think that some structure could help us a lot” (interviewee)
- There is no ethical tool that has the aim to formulate policies or laws with yet.

B) Lack of responsibility allocation

- Taking part in the ethical discussion was perceived as unnecessary by one of the interviewed startup founders. “We arrange the technology, it is not up to us to determine how it is used. From a technical point of view, we look at how privacy can be guaranteed (solutions), while the ethical discussion revolves more around how we want to solve it (opportunities).” (interviewee)
- “For ethical tools I would look at examples in the market and how people respond to them to determine if my ethical compass is correct. Otherwise I just deliver to the customer, I formulate it in such a way that they are responsible”. (interviewee)
- “They do it their own way anyways” (interviewee). Hence, it emphasizes the importance that people acknowledge the needed responsibility and involvement.
- “Current tools are made for AI developers (technical viewpoint), designers (social viewpoint) or for managers (legal or ethical viewpoint), not to collaboratively filling it in with multiple perspectives. (...) Those impact assessments are done internally, not including other parties. There is a demand for an inclusive societal debate with experts that cover the topics of ethics, legal, technical and social viewpoints. (...) As soon as you bring in a lawyer, you find out that the limiting factor is not the law but the unclarity. It is assumed that it is not allowed anyway.” (interviewee)

C) Tools are not inviting

- Current tools provide the feeling of “checking of boxes” (interviewee), it is not perceived as an inviting or fun experience.
- Filling in an ethical tool is experienced as a side activity, not part of the process.
- The word ‘ethics’ is associated with a heavy load of discussion.

D) The language barrier

- “There are three languages in the public-private collaboration” (interviewee). This is considered a challenge when gathering these representatives in the room. Nowadays law enforcement is often a matter of multidisciplinary cooperation, involving public administration, justice and private parties. They work together in chains and networks. Three disciplines are involved in almost every partnership: criminal and sanction law (because that is the hard core of enforcement), chain-computerization (because information must be exchanged in and between chains) and data protection (privacy). Borst (2019) offers a way to define these three perspectives in their mutual relationship. Table 2 (on the next page) represents these perspectives, created based upon the interview and the book reference.

Table 2: The three languages of public administration, justice and private parties

Role	Terminology	Language
Legislator	Raison d'être	Language of Politics (French)
Policymaker / Manager	Targets, Mission statements, Planning & control, IT, Balanced Score Cards	Language of Scientific Management (American English)
Executor (Police, service)	Punktlichkeit, Gründlichkeit, the language of Weber and bureaucracy	Language of precision, consistency, predictability, rules and processes (German)

The convergence and concreteness of ethical dialogue

This language can be bridged by using examples, visual references or stories people can empathize with. An approach described by Verbeek & Tijink (2019) is the 'guided ethics' approach (see Figure 37). This approach helps to weigh out the policy options starting the dialogue from specific cases regarding the technology. Referring to specific cases helps to imagine concrete results, this way you avoid that the discussion stays theoretical and abstract. By adding this third reference point, you make ideas easier to communicate and bridges gaps in communication. Therefore, using experiment cases as inspiring projects help you better imagine how desirable such AI applications are and how it could work (against you).

A FRAMEWORK FOR GUIDED ETHICS

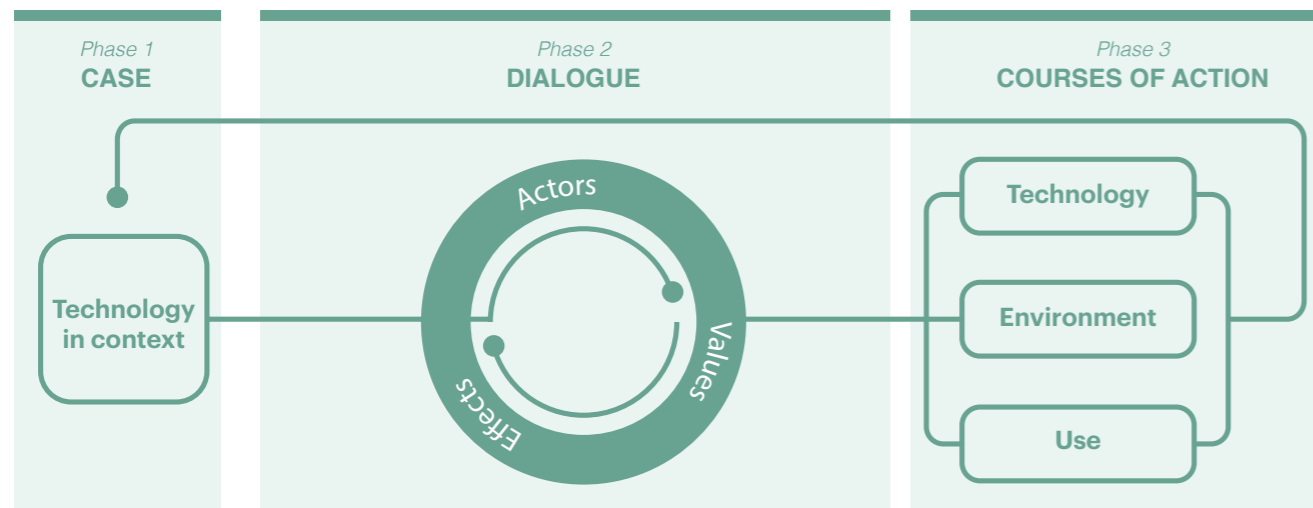


Figure 37: The Guided Ethics framework (Verbeek & Tijink, 2019)

2.3.4. Insights & Conclusion

THE BUILDING BLOCKS



Empathy:

Much attention is currently paid to the role of ethics and human rights in the development and roll-out of AI applications. There are also many public and private initiatives which focus on creating standards and frameworks for AI. What is missing, is convergence and concreteness of all these relevant and important actors. Aspects such as inclusiveness, privacy and equal treatment have its own flavour based on angle or perspective one' looks from. To understand what future we want to achieve, it's crucial to clarify what we as society desire, in order to carve a path towards that future.



Inclusive:

Governmental innovation requires an additional 'lens' for allowability, building upon the sweet spot of business innovation representing the lenses of: desirability, feasibility and viability. Current approach is focused on learning what's desirable by knowing what's feasible (= technology push), while we should first define what's ought to be desirable (=value-driven), then craving the path towards what's feasible. In order to successfully balance these lenses, a collaboration with the representation of these domains (government, society, academia and business) is proposed. As the current approach is focused on learning what's desirable by knowing what's feasible (= technology push), we should switch the order to first define what's ought to be desirable (=value-driven), then craving the path towards what's feasible. Inviting other perspectives helps to understand others point of views and different tastes on ethics of an AI application. Having conflicting professional values working on the same task can bring friction. Therefore, it is crucial to find overlap and reach alignment in the goal. Work towards consensus.



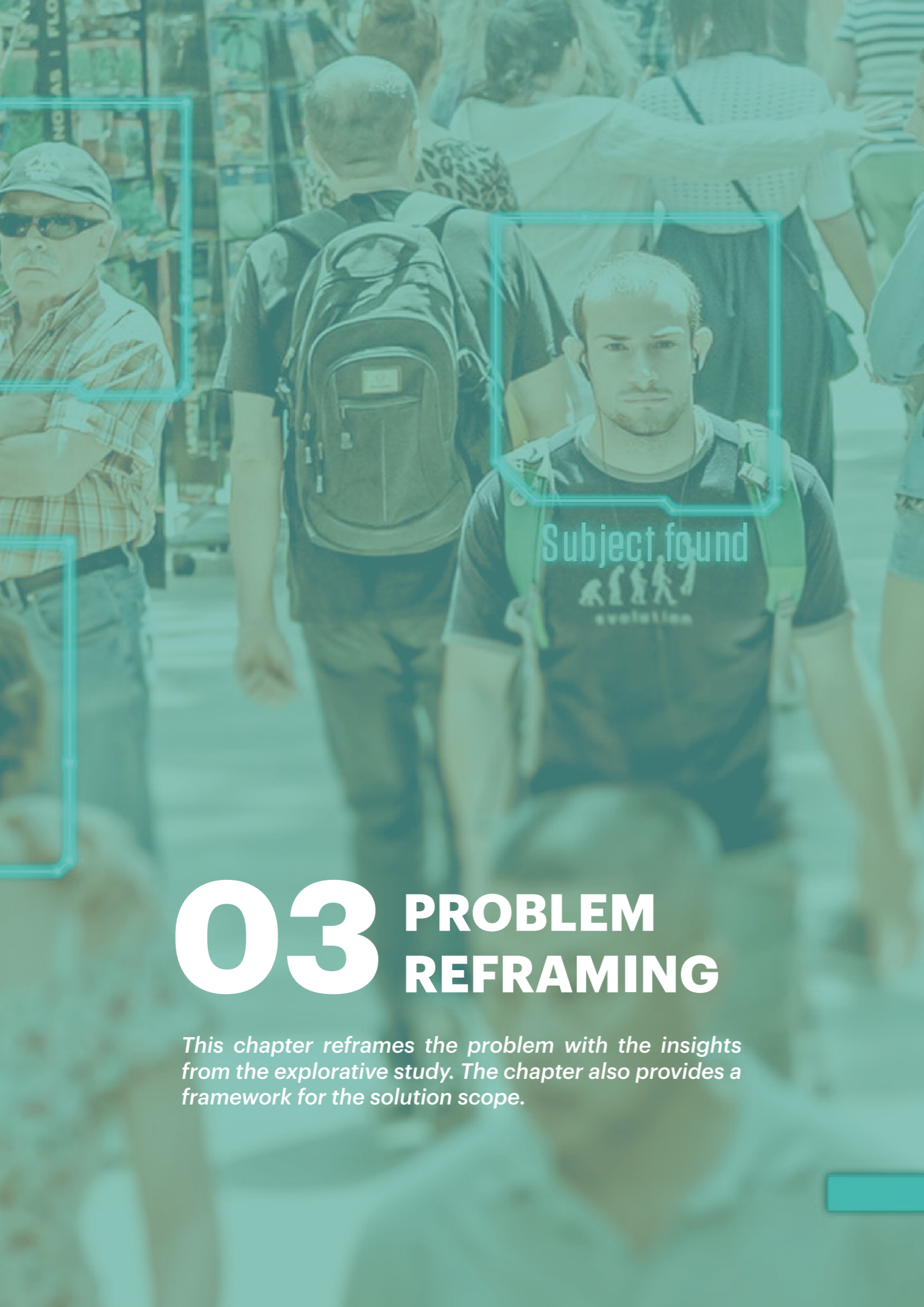
Reflective:

Let failure enable revelation. "Faced with the choice between changing one's mind and proving that there is no need to do so, almost everyone gets busy on the proof". Meaning, it is easier to continue to work with what we are used to than to change our minds and accept and learn new ways of dealing with these wicked problems. When an experiment fails, it is seen as a negative experience, while in essence it is an opportunity to distract valuable learnings.



Demonstrative:

Starting the dialogue using technology in the context, to avoid theoretical dialogue and demystify the desirability of AI applications. Use actual ongoing experiments as tangible references. Visuals and storytelling bridge the language barrier of other fields of expertise.



03 PROBLEM REFRAMING

This chapter reframes the problem with the insights from the explorative study. The chapter also provides a framework for the solution scope.

3.1 Revealing the key problem

Previous chapters lead to the gained insights for the design space. A clustering lead to addressing the key problem and framing the problem statement.

3.1.1. Clustered problem landscape

The clustered problem landscape represents the three U's of Uncertainty: Unpredictability, unclarity and unfamiliarity. Figure 38 visualizes the clusters of the key problem areas:

- the *unpredictability* of AI: resulting in a shift of the no-failure culture to an open mindset.
- the *unclearly defined desirability* by society, resulting in the need for stakeholder inclusion and a tool to demystify the fuzziness of what the public defines 'ethical' and desirable.
- the *unfamiliarity* of coping with new technology: resulting in the need to shift to organizational learning, meaning involving IT expertise in decision-making, and learn through experimenting.

As these sources of uncertainty all take place within the rapid pace of tech development, it emphasizes the need for an agile approach to policymaking.

PROBLEM LANDSCAPE IN RAPID PACE OF TECH DEVELOPMENT

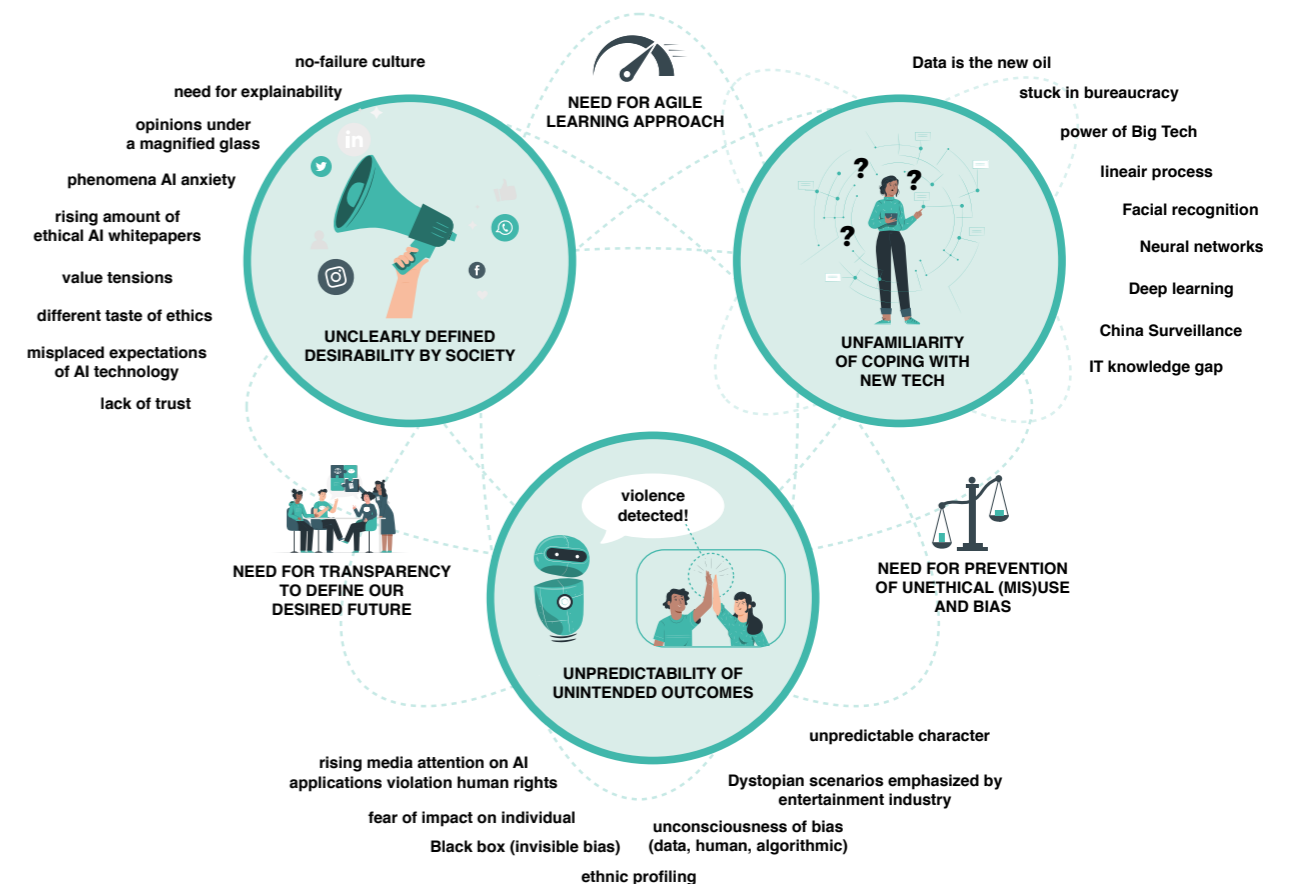


Figure 38: Representation of the clustered problem areas in the problem landscape

3.1.2 Revised problem statement

Policymakers struggle with formulating policies steering towards human-centered AI development because they are under extreme pressure by the unclearly defined desirability by society, the unfamiliarity of coping with the new technology and the unpredictability of the consequences caused by it, all happening within the rapid pace of tech development.

3.2 A framework for agile policymaking

Combine the 5 steps of Design Thinking, the lenses of governmental innovation, and the guided ethics framework. The sum of these elements result in a framework for Agile policymaking, visualized in Figure 39.

The chronological order to start with a framed context, being a concrete case example as a tangible reference to depart the dialogue from. Then empathise to define what is desired by the representative stakeholders, ideate to understand how the needs can be embedded in the system and build up on each others knowledge, prototype to learn about the feasibility rapidly, and test the new policy and regulations to validate assumptions.

Therefore, a structural shift in the mindset and organization is needed. A manifesto can be seen as a document publicly declaring the position or program of its issuer. It advances a set of ideas, opinions, or views, but it can also lay out a plan of action. It can address any topic and is generally written in the name of a group sharing a common perspective, ideology, or purpose rather than in an individual. Therefore, to activate this shift, it requires a manifesto explaining the new mindset.

FRAMEWORK FOR AGILE POLICYMAKING

Guided Ethics

+

Design Thinking

+

Agile Sprints

PEOPLE:



PROCESS:

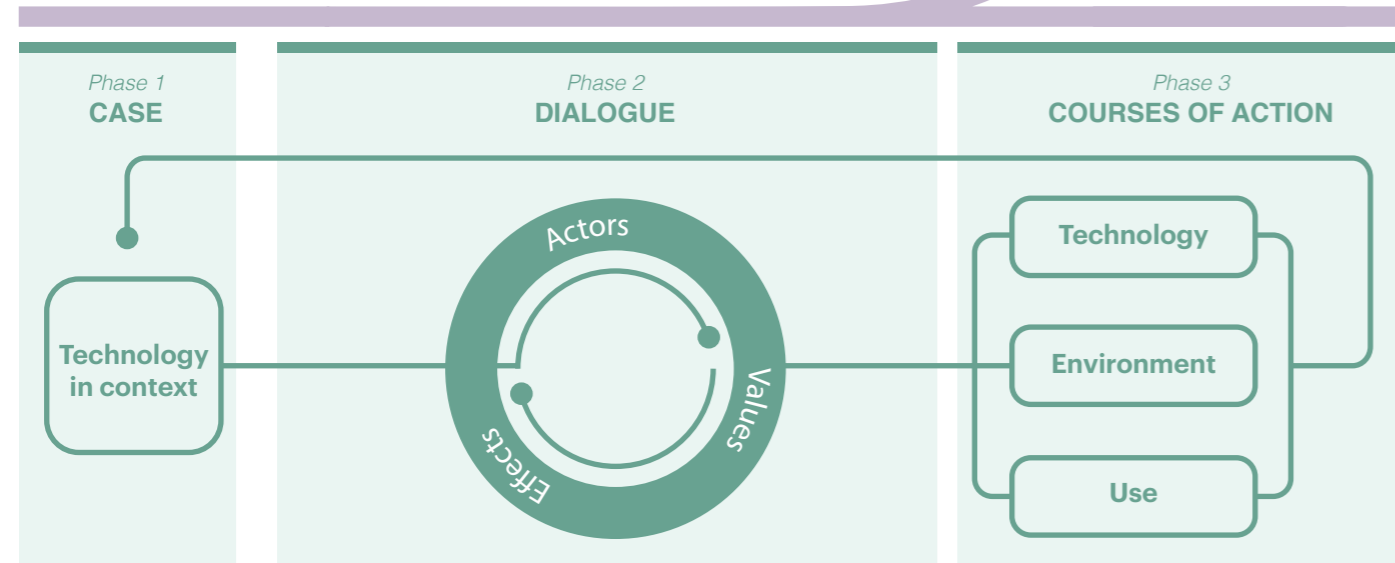
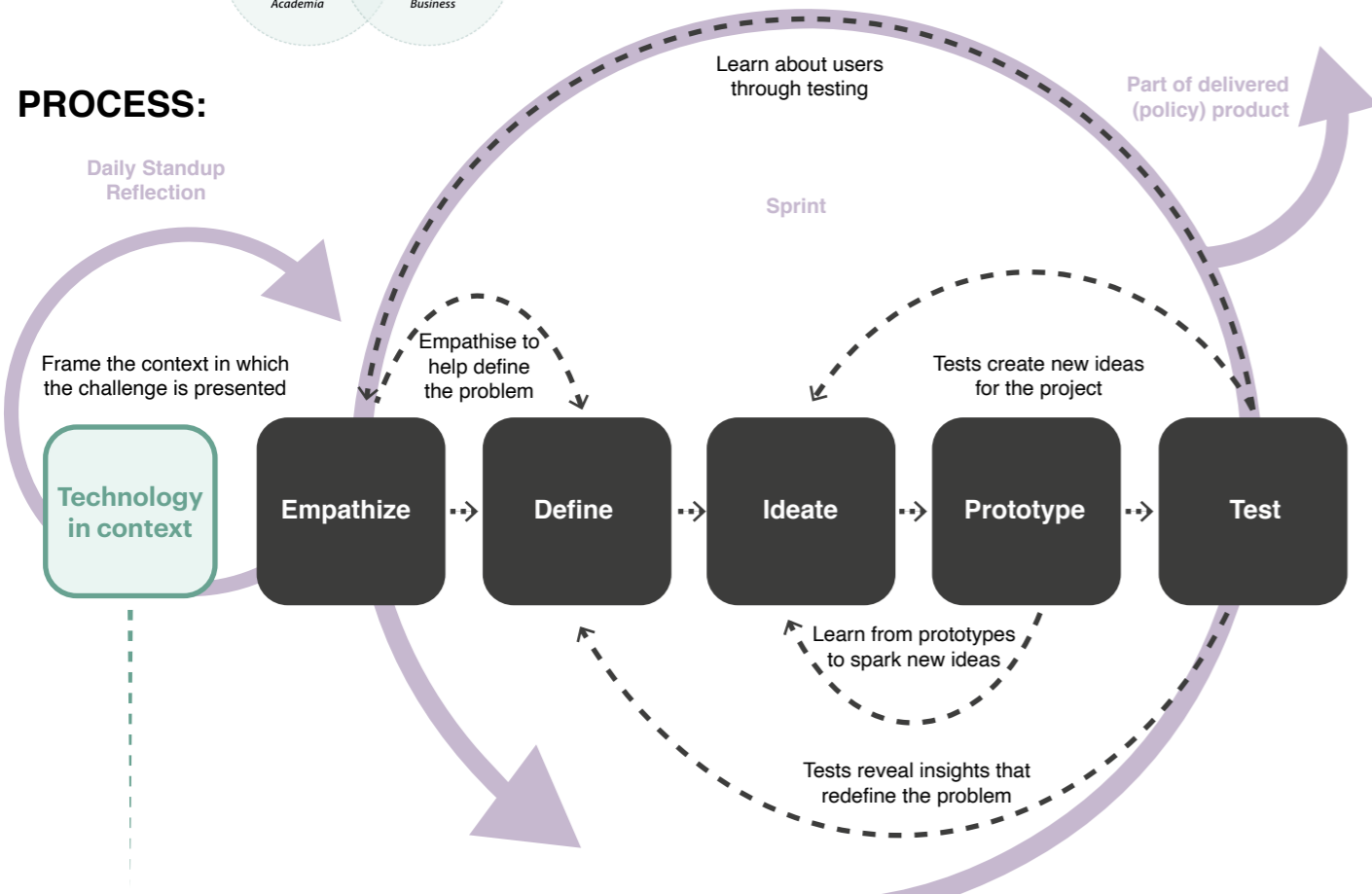


Figure 39: A framework for Agile Policymaking



04 DESIGN VISION

This chapter introduces the design vision and corresponding manifesto. In addition, the requirements of the physical design are given that will make the design vision and manifesto come to life.

4.1 Design vision (the why)

The ministry should frame experiments as a means to facilitate the dialogue on the desirability of AI system development.

Design vision: (AI) experiments are a priority vehicle for organisational learning

The framework for agile policymaking is applied to this design vision. Figure 40 and Figure 41 describe the shift from current to the desired way of working, respondingly.

The proactive policymaking process is recognized as a future-oriented exploration. Experiments are used to gain feedback early in the process. Policy Prototypes and Minimum Viable Policies are created as first feedback loops prior to current feasibility studies. An inclusive panel, represented by the Quadruple Helix domains reduce the blind spots and facilitate the short-line communication and alignment from the relevant stakeholders. The potential interventions transcend theoretical discussion as the conducted experiments collect proof and provide a context for detailed dialogue. The daily activities of a policymaker will not be swayed by the issues of the day anymore (i.e. a slip of the tongue of a minister). Hence, the role of the policymaker is less affected by the magnifying glass of the media.

REACTIVE POLICYMAKING: SWAYED BY THE ISSUES OF THE DAY

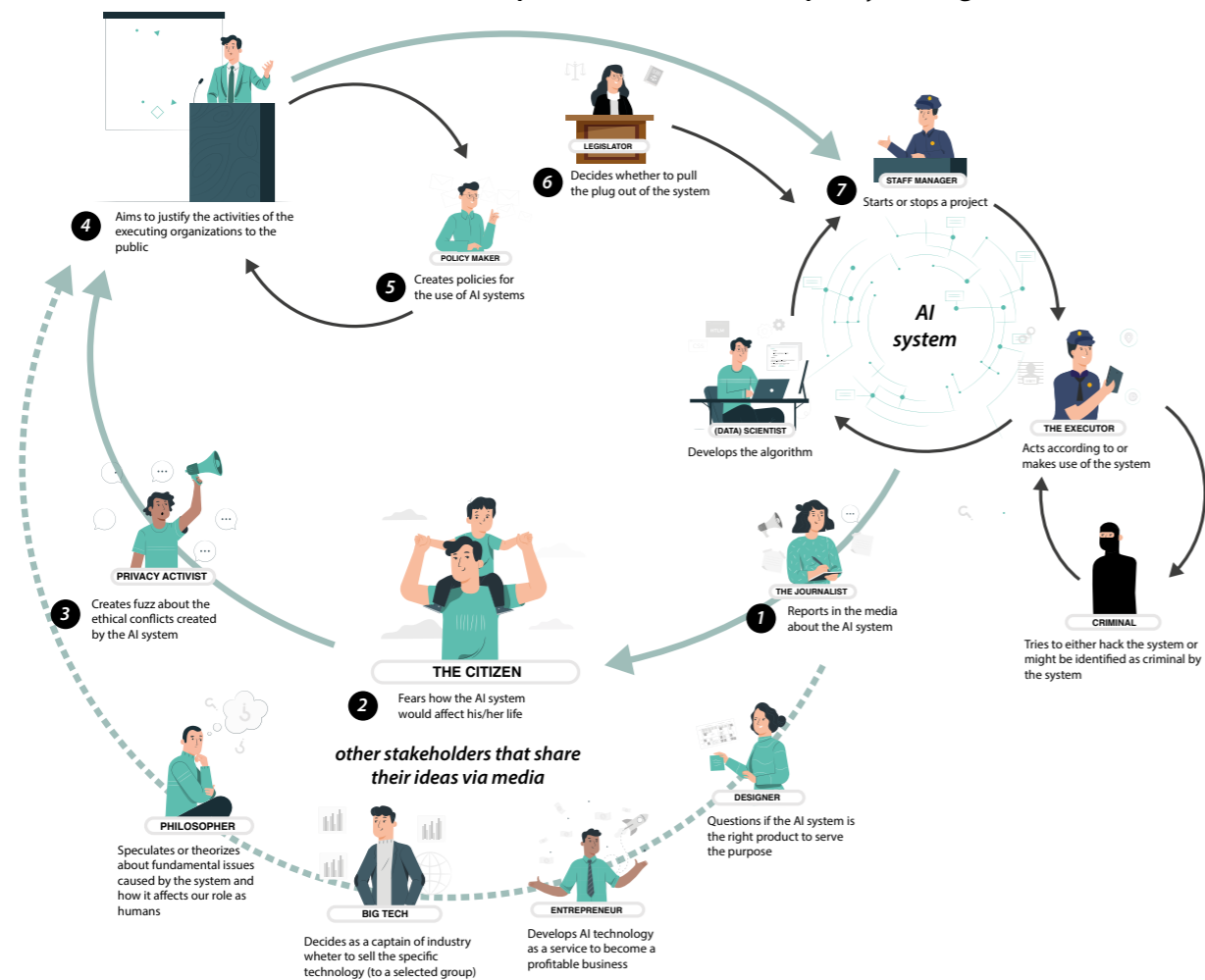


Figure 40: reactive policymaking caused by the magnifying glass of the media

PROACTIVE POLICYMAKING: A FUTURE-ORIENTED EXPLORATION

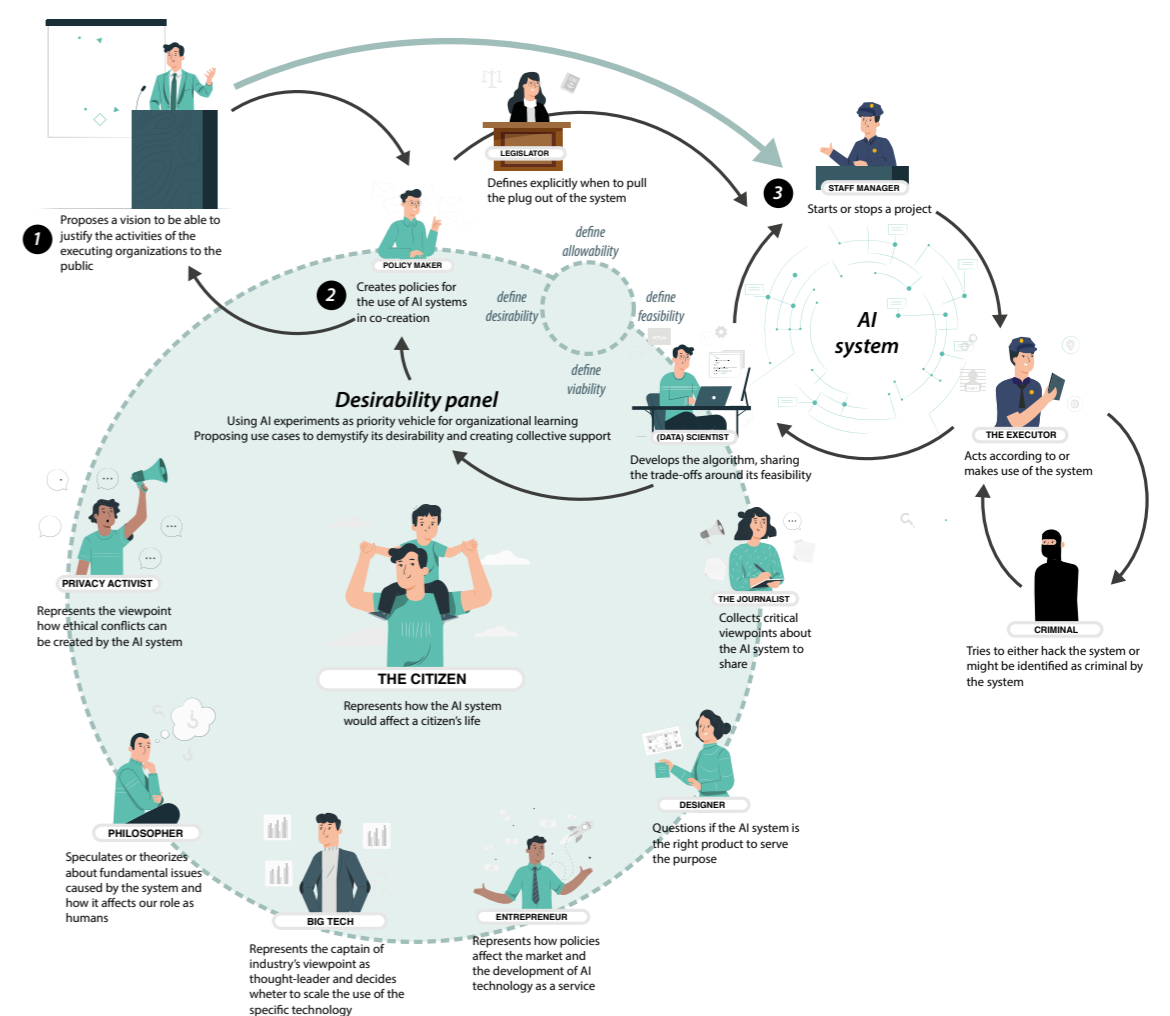


Figure 40: A future-oriented proactive, inclusive and agile representation of policymaking

4.2 CERTIFIED manifesto (the how)

To facilitate the transition management, leadership starts with a recognizable manifesto.

The collected insights from the explorative study resemble the building blocks towards the design vision, introducing: the CERTIFIED approach towards human-centered AI adoption. CERTIFIED is the acronym for: Creative, Empathic, Reflective, Transparent, Inclusive, Future-oriented, Iterative, Experience-driven, and Demonstrative. The manifesto is shown in Figure 42 on the next page. The Dutch version can be found in Appendix D.

Our Manifesto

*We believe in a **CERTIFIED** approach towards facilitating human-centered AI adoption. We cope with wicked problems with lots of uncertainties, where there is no path that leads us the way. Due to the adaptive nature of AI and the fact that society's needs change over time, we have to learn by doing to understand what services we need to offer within the domain of Safety and Justice. Therefore we need to shape our own path, inspired by the agile methodology. Innovation is our duty, positioning the human in the center of technology is our goal.*



*We generate innovative solutions, empowered by **Creativity**. We invite for play, with an open and positive mindset, to come up with new ideas in co-creation.*



*We avoid reactive policy making and being swayed by the issues of the day. We work vision-driven, spending structural time on **Future-orientation**. We aspire to inspire, using our vision as foundation for short- and longterm decisions.*



*We generate support, starting with **Empathizing** with the affected parties. We examine latent needs by including various viewpoints on the context, before deciding to put our ideas in practice.*



*We provide a safe environment for potential failure and bring **Iteration** into the policy development process. We deliver intermediate results in short cycles by introducing 'Minimum Viable Policies' and test contextual micro-rules in parallel.*



*We generate learning capacity, keeping an eye on both process and result, by **Reflecting** on the system as a whole. We embrace criticism, as a complaint is a gift, containing rich insights of experiences.*



*To obtain the know-how, our choices are **Experience-driven**, with structural expert panels, defining policies with the 'whole system in the room'. We validate our assumptions with 'experience panels' that represent the Quadruple helix.*



*Comprehension is achieved by our **Transparency**. Our acts need to be understandable for all. Thus, taking people along in our thoughts means being open about our experiment processes, offering a low entrance barrier to co-develop better outcomes.*



*We give meaning not only by words. We refer to experiment cases to facilitate rich discussions that are tangible. We bridge languages by **Demonstrations**, framing experiments as a means to facilitate societal dialogue on what future we desire, bringing transparency to the public.*



*We value **Inclusiveness**, powered by open access to dialogue at the start of experiments. We empower face-to-face alignment in vertical, horizontal, cross-silo, and interdepartmental teams, lowering the barrier between policy and execution. This will lead to less resistance, ignorance and incomprehension by the public and within the organisation.*

4.3 Toolkit Design (the what)

The goal of a design artefact is to make the CERTIFIED approach come to life. A way is to provide an experience to live the approach itself. To offer this experience, a toolkit is created to guide the first step towards agile policymaking.

4.3.1. Reasoning for the creation of a toolkit

Currently, there is the lack of an action perspective for the created manifesto. A toolkit could offer a standardized guided approach to take the first step towards change. People need to understand what actions need to be considered by living the experience of how CERTIFIED can suit their daily practices. A toolkit will also improve the awareness of the new approach with a tangible shareable artefact. Research on the limitations of current ethical tools showed that a 'context' is needed in order to avoid the dialogue from being too generic or universal. To conclude, in current tools, the different perspectives are missing

Hence, the design challenge for the toolkit is formulated as:

'Design a normative practical guiding tool for agile policymaking, offering policymakers a low-barrier, hands-on action perspective to apply the CERTIFIED approach into their daily practices.'

4.3.2 Toolkit design requirements

The requirements for the toolkit follow the lenses of business innovation by Calabretta (2016): feasibility, viability, and desirability. These requirements are set as frame for the ideation.

Feasibility:

- The toolkit offers the ability to collaborate with a variety of stakeholders from remote (due to Covid pandemic), offering the potential of using the toolkit online
- The toolkit offers the ability to empathise with stakeholders with a design artefact
- The toolkit and needed software have a low entrance-barrier, meaning no prior knowledge or additional steps to require or purchase components

Viability:

- The toolkit guides a structural agile approach, using canvases to guide thought exercises and safeguard iteration in the process.
- The toolkit incites to act, guiding the creation and testing of minimum viable policies for validation
- The toolkit supports the creation of Human-Centered AI using AI experiments as references to define policies
- The toolkit facilitates knowledge building my means of eliciting new information

Desirability:

- the toolkit is inviting to use as the appearance is attractive
- the toolkit offers a fun experience, including an 'invite for play' element
- the toolkit has the wow-factor, introducing a new way of working
- the toolkit provides a feeling of control, offering the ability to act fast and adaptive



5.1 Conceptualisation

5.1.1. Ideation

An online creative session with three student groups was held to fuel the ideation process, represented in Figure 43. Students were asked to ideate on how to offer an inclusive and transparent service as government on the use of AI without the ability of criminals to abuse this service.

ONLINE IDEATION TO FUEL INSPIRATION

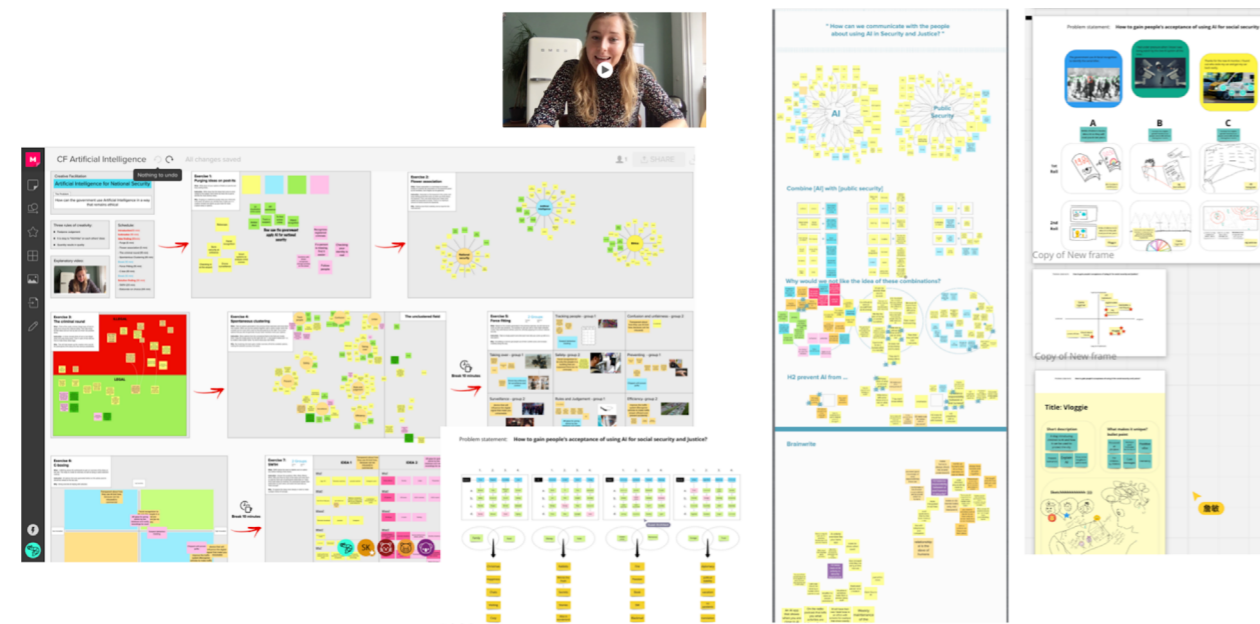


Figure 43: The facilitated ideation in MURAL

Some of the ideas included: a talk show “ConversAltions”, TV commercial on explaining how transparency is met, an app for children to learn about how AI affects your daily life, or AR glasses that provide real-time law advice. The main insight was that a few concepts addressed potential abuse by only sending information (TV commercial), while the preferred option was to seek interaction and thus input (talkshow, app). Most ideas emphasized the tone of voice and language as a crucial part in the communication with ‘outsiders’. Inclusivity was addressed by reaching out to a variety of people.

5.1.2. Iterative process

The establishment of the final concept covered an iterative process with various ideas, thinking prototypes, interviews to validate the general idea of the toolkit, observation tests on the use of the toolkit and finally evaluations with the users to discuss if the previously stated requirements have been met. Tests were held with both the physical tool as well as the digital tool. Figure 44 shows the process from several points of time in the conceptualisation phase.

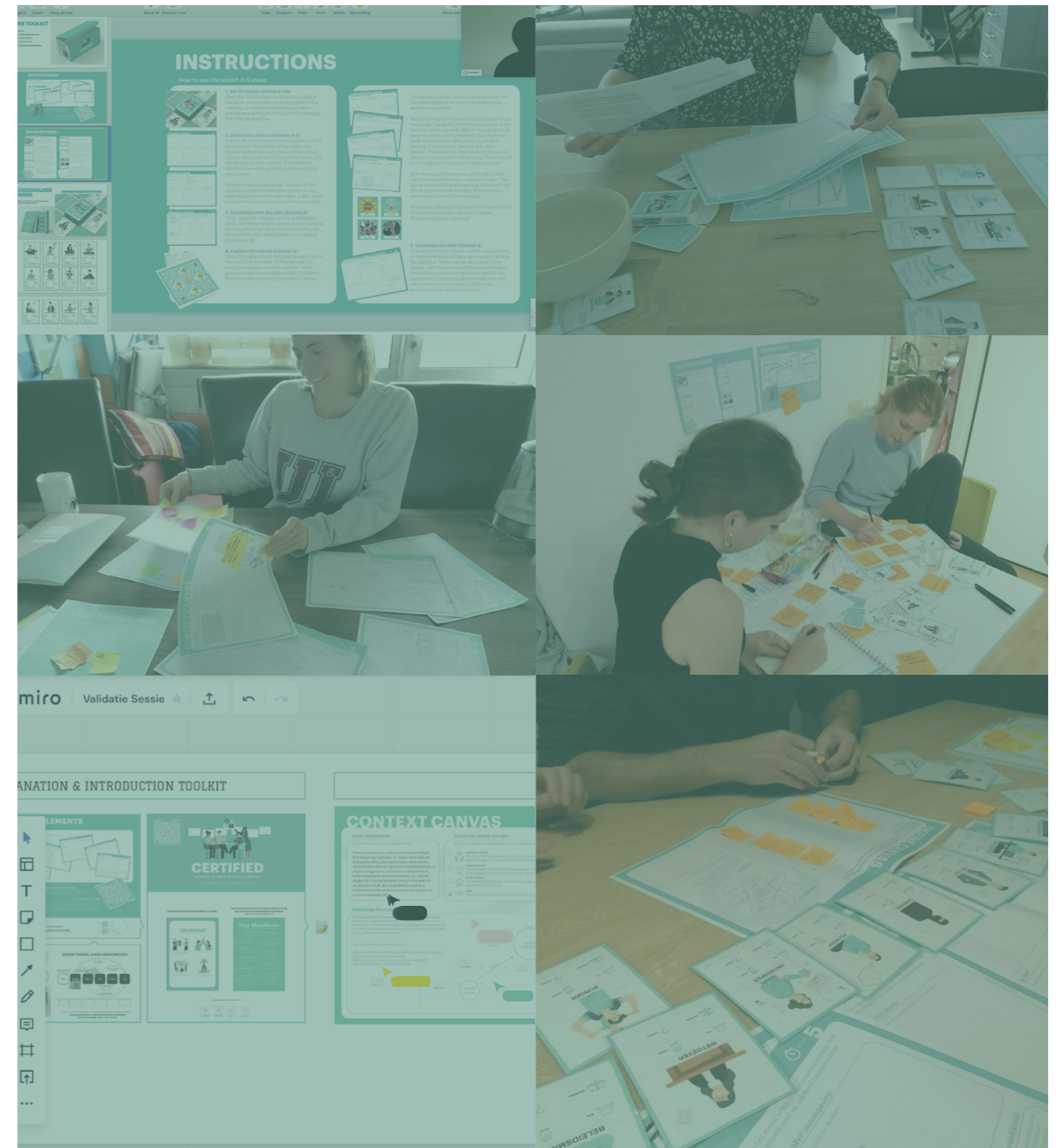


Figure 44: Various images from the iterative design process of both physical and digital toolkit..

5.2 Final concept

After an iterative process, the final concept is created. The final concept consists of various elements that support the framework, represented in Figure 45:

THE CERTIFIED APPROACH

FRAMEWORK FOR AGILE POLICYMAKING



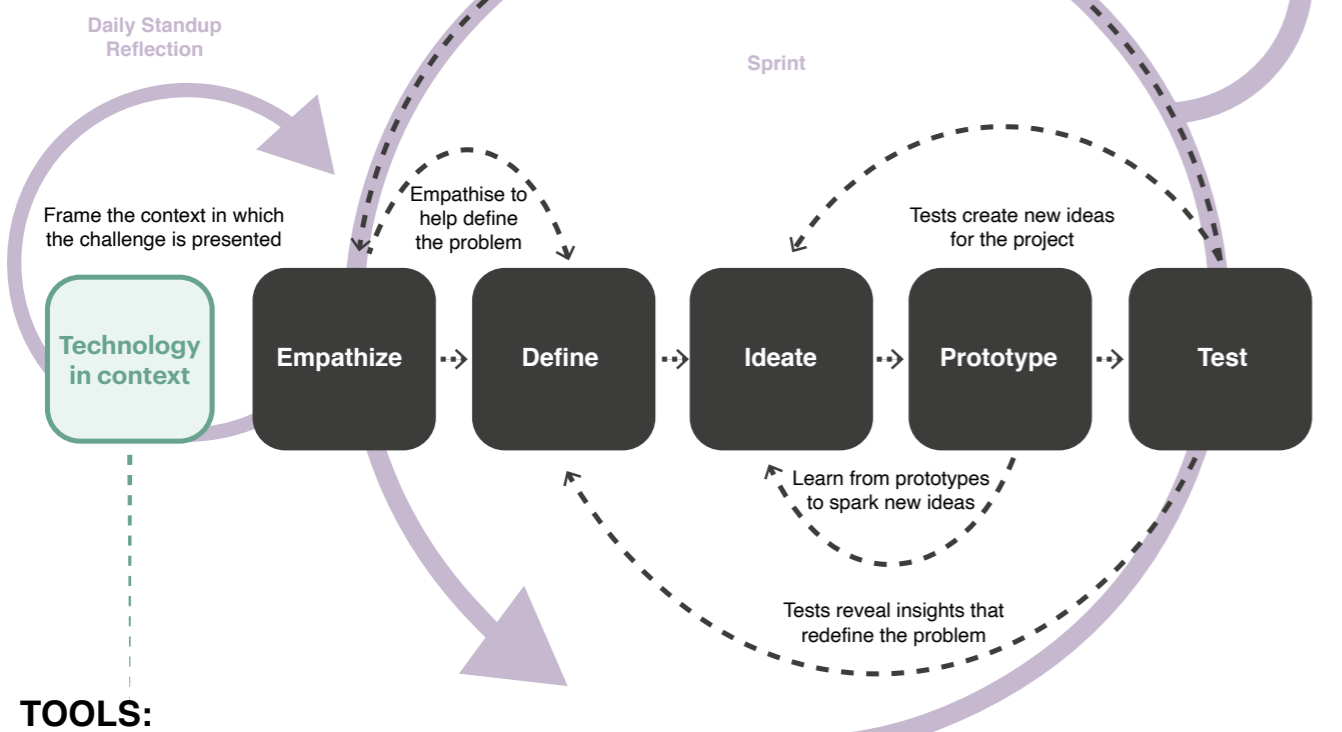
MINDSET:



PEOPLE:



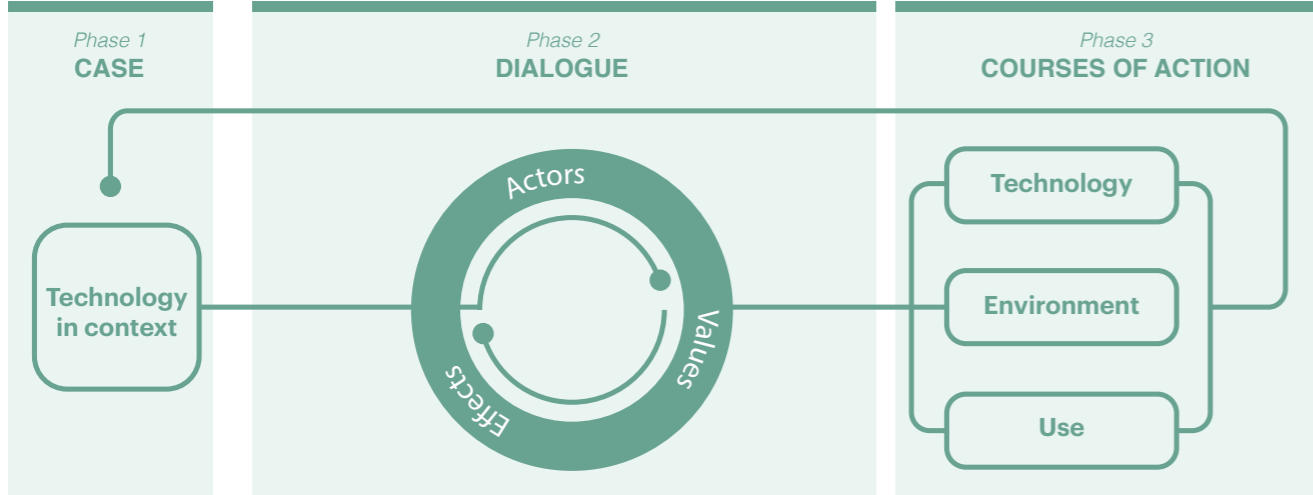
PROCESS:



TOOLS:



CONTENT:



TOOLKIT ELEMENTS

The toolkit contains:

De process canvases:

These canvases structure the use of this toolkit and capture the insights from the dialogue.



Trigger cards:

These cards introduce questions to stimulate out-of-the-box thinking



Understanding Perspectives Card deck:

This card deck is used to include multiple perspectives into the act of formulating policies for a given context.



Tech-Innovation Compass Poster:

This visual representation is used as inspiring stimuli in the room to stimulate an exploratory mindset. In doing so, the metaphor helps to explore the lenses of *desirability*, *viability* and *feasibility* to determine the *allowability* level.

Figure 45: A representation of how the toolkit elements link to the framework

Figure 46: The toolkit elements

5.2.1. The CERTIFIED toolkit

The toolkit invites us to start the dialogue on what is found desirable by society, instead of looking at what's feasible. This goes beyond asking people what they want, but let a group experience the borders of what is found desirable by real-case examples. By collecting these insights in the toolkit, patterns will emerge and the debate on the desirability regarding the policy intervention for AI can be made more explicit. This offers the possibility to observe, rather than solely ask, referring to a famous quote by Ford: *"If we'd ask the people, they would have wanted faster horses"*.

The toolkit will help you to:

- understand the principles of design thinking and how they can be used in a government context
- explore the different stages of the design process, and why it is important to place citizens and users at the centre of your work
- start using new practical tools and methods to create change in how you tackle problems, both old and new
- consider the challenges you might come across when trying to implement this new approach, and how you might overcome them.

The toolkit includes a perspective card deck, a set of canvases, and a poster of a compass, being a visual reference to the Governmental Innovation lenses. The toolkit elements are shown in Figure 46.

Perspectives Card Deck

The idea draws upon the Six Thinking Hats (De Bono, 1986), as it can be seen as a role-playing model for team-based problem solving. In this case, the Perspectives Card Deck is used as a brainstorming technique to explore problems and solutions and uncover ideas that might otherwise be overlooked by a homogeneously thinking group. Therefore, these cards are used to gain empathy from the multiple perspectives of society, in order to formulate a thorough policy for the given context.

The cards introduce a stakeholder, the roles, actions, values and power related to the experiment, and questions that derive from this perspective. The carddeck is visualized in Figure 47. The cards are created, iterated and validated with representative stakeholders through online screen-sharing validation sessions. The final carddeck can be found in Appendix E.

The card deck includes a few blanco cards, so the parties who aren't represented in the cards can be filled in. The cards can be used in 3 different ways:

1. Choose and evaluate your perspective at the start of the experience panel meeting.
2. Choose a perspective not associated with your expertise, and use this perspective throughout the experience panel meeting.
3. Use the cards with your own team to empathize with the other parties to evaluate your team's performances and ideas



Figures 47: The Perspectives Card deck

The added value of the card deck:

- Learn you can't do ethics alone. Such a social issue requires several perspectives.
- By experiencing it from someone else's perspective, you identify your blind spots
- Incite to take action: to go to the streets more often and check your assumptions with the public
- In case it is not possible to collect a group of representatives of society, an unrepresented stakeholder can still be visually represented during the dialogue.

5.2.2. The setting

Participants

The toolkit will be used by policymakers. They represent the Problem Owner to formulate policies on new technologies. An external moderator is requested to facilitate the toolkit in the session. The Problem Owner will organize a session with the Desirability panel to collaboratively use the canvasses as structure for the dialogue. The Desirability panel includes a group of representatives of the four areas (government, business, society and academia), inspired by the sweet spot of governmental innovation (see Figure 51 for the similarity).

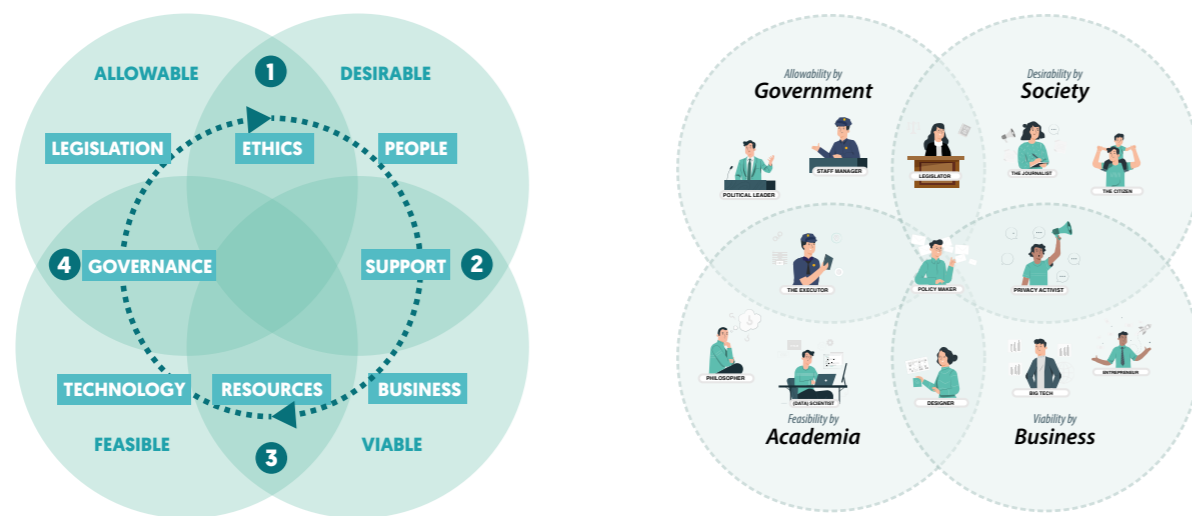


Figure 51: The similarity of the theory behind the sweet spot of governmental innovation and the Desirability Panel.

Occasion

During interviews with several policymakers, it became clear that the toolkit is most suitable for the policymaker on four specified occasions, depicted in Figure 52. The figure also resembles how much time is taken for these occasions and how much time is expected to be considered in the execution of the toolkit (elements). The minimal timeframe recommended for a complete run through the toolkit is a two-day sprint. The time to go through the toolkit is 4-5 hours at least, but it is advised to introduce breaks after each canvas and/or spreading the toolkit over several days. An optimal timeframe would be spending 2 hours a week for 1 canvas, to provide more time for reflection.

TOOLKIT USE IN RELATION TO OCCASIONS

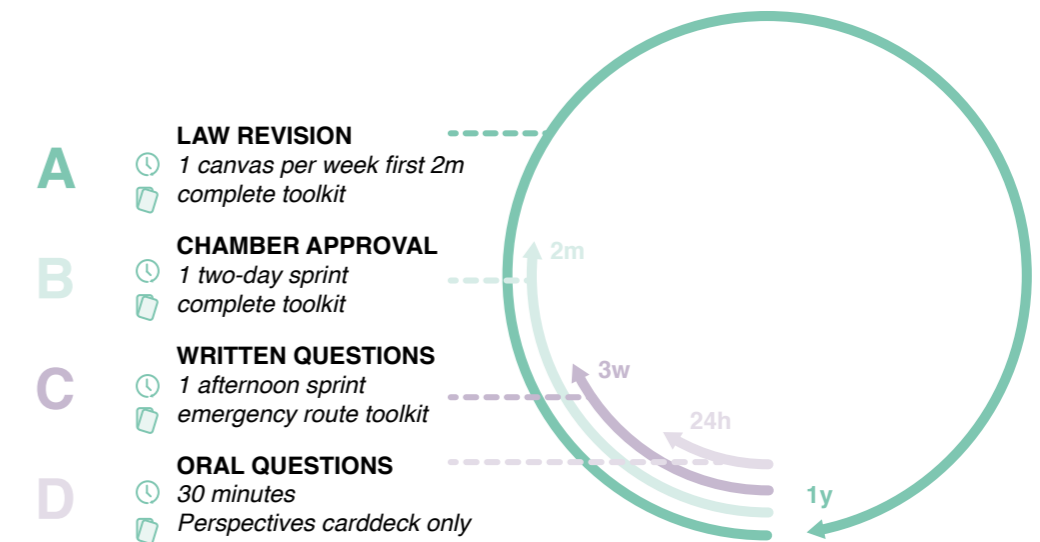


Figure 52: The four action timelines of a policymaker

The four occasions are:

Law revision:

This average process timeline is 1 year. To revise a law, a complete run through the canvasses would provide sufficient insights as a kick-start for the CERTIFIED mindset. The first 2 months, 1 canvas a week is completed, to continuously reflect and build upon the gathered insights early in the process. It provides enough space to invite additional stakeholders for each canvas and validate the stated assumptions. For example: After completing the Empathy canvas, the assumptions of the chosen perspectives can be validated by interviewing these representatives, if this perspective is not represented in the panel. After indicating the (potential) blind spots, the following canvas builds upon these insights.

Chamber approval:

This average process timeline is 2 months. Therefore, a kick-start with a 2-day sprint is sufficient to go through the complete toolkit with enough room for reflection.

Written question:

The average process timeline for answering a written question is 3 weeks. As the toolkit is a new instrument, it is likely that the time pressure during the first pilots for a written question will not leave much room for completing the entire toolkit. Therefore, an emergency route with a time-bounded 1-day sprint (or pressure cooker) is advised.

Oral question:

The average process timeline for answering an oral question is usually less than 24 hours. This requires a quick response. One interviewee explained it would be too much time to focus on anything else rather than typing and 'getting on the line'. Hence, a quick view over the perspective card deck would be sufficient to consider the relevant stakeholders.

Preparation

All participants should be sent appropriate briefing documents at least two weeks in advance of the meeting. These documents are produced and distributed to ensure that all participants can familiarise themselves with the methodology and the program software MIRO before the online meeting is hosted. The participant invitation and briefing documents should include the following:

- workshop aims;
- general introduction, e.g. impacts of agile policymaking;
- need for participatory tools;
- role of the toolkit in relation to the aim of the CERTIFIED manifesto
- the agenda invite
- the link to the MIRO board

5.2.3. The implementation

In order to structurally embed the CERTIFIED approach in the organization, the AIDA steps are considered (Awareness, Interest, Desire, Action) in relation to the stages of the Technology adoption Lifecycle (Bohlen & Beal, 1957), see Figure 53.

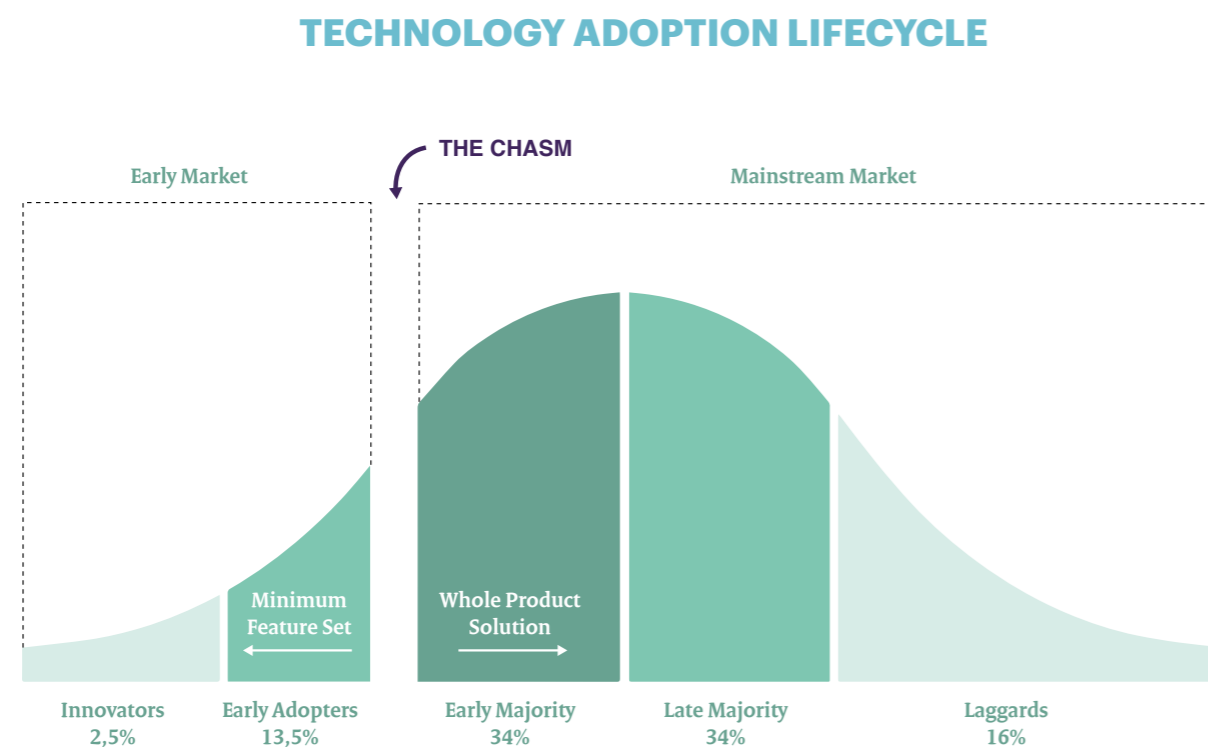


Figure 53: The stages of the technology adoption Lifecycle (Bohlen & Beal, 1957)

Awareness

First, the CERTIFIED manifesto and the toolkit need to be spread throughout the organization in order to create awareness of its presence. To draw the attention of policymakers, information should be spread throughout their commonly used channels such as LinkedIn and the ministry's intranet. Presence at webinars and symposia organized by the government help gain traction of the acronym. To smoothen the introduction of the CERTIFIED manifesto, it would be helpful to create an animation or short clip with visual references that reaches beyond the use of icons. This could be easily shared and improves the viewer's understanding. It also lowers the barrier as it takes less effort than reading the textual version.

Interest

The second step is to arouse interest by pointing out the added value of the toolkit in their current way of working. The best way to do this is to offer them to experience the toolkit with a 'Certified' moderator. The government is a network culture, meaning you can solely focus on providing the policymakers, who acknowledge the need to change, an attractive offer which they cannot resist. Creating a low barrier to taste what it is like to use the toolkit can be achieved by voluntarily hosting a webinar, or proactively reaching out to policymakers offering them the manifesto as the solution. Reading the manifesto should awaken the desire to act accordingly and become a 'Certified ambassador'. If the response is: Yes, but how? ", offer a session to walk through the toolkit. Develop tools that can easily be shared like a short animation or video explaining the reasoning behind the manifesto. As these early adopters can empathize with the need, it could spread as wildfire throughout the organization. This will also gain attention by the ones higher up, who have the mandate. One side note: use the jargon familiar by policymakers to prevent misunderstanding.

Desire

Convince the board in a meeting by showing the promised land, the identified obstacles and how to overcome them. Also visually present what would happen and what the ministry would miss out on if they won't take action now. Stress it's their responsibility and offer help. When the urgency is cultivated, time and budget will be obtained to organize sessions.

Action

As a starting point, pilots will be held upon request with the AI team as moderators, as they have been actively involved in the development of the toolkit. Later on, this task will be allocated to the 'Certified moderators', as the crucial part of the toolkit is the fact that it needs a good moderator. The timekeeping and ability to bridge the languages between the different parties has a great effect on the process. These moderators help to scale the distribution of the toolkit and preserve the quality. Certified moderators can publicly update their status, arousing curiosity by their colleagues. Let colleagues express the use of the toolkit as the preferred and innovative way of working to become an adaptive and resilient government.

Train the trainer: Facilitator role

Research on the use of the DEDA framework by Van den Hoven (2020) indicated that processes with a moderator bring the most fruitful outcomes and have therefore decided to make it obligatory to run through the DEDA framework with a moderator first. After answering all the questions on the DEDA framework, and once it is clear what next steps need to be taken, a moderator is no longer necessary. The Utrecht Data School offers “train the trainer” workshops to introduce team members into the facilitation process and prepare them for this role. This process is also recommended as part of maintaining the quality of the manifesto and its execution by means of the toolkit. Certified moderators can acquire a Certified Certificate after passing the train-the-trainer workshop to become a Certified Trainer.

Online vs offline

Due to the pandemic, it is not possible to go through the toolkit together physically. When the pandemic has come to an end, the toolkit could also be used in its physical state. A CERTIFIED box contains the Perspectives card deck, the Manifesto on A4, the canvases on A3, a set of post-its and the Compass poster as a visual reminder. The physical toolkit is visualized in Figure 54.

THE PHYSICAL CERTIFIED TOOLKIT INCLUDING ELEMENTS

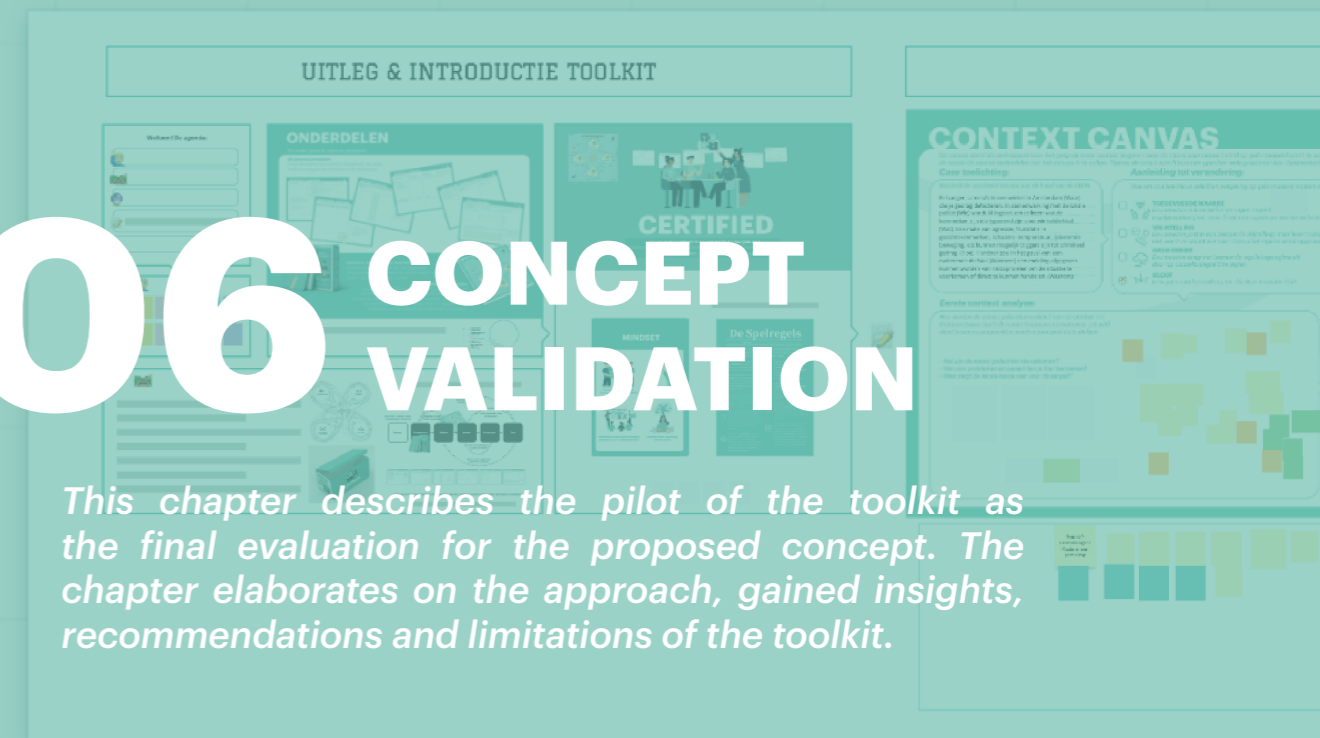


Figure 54: The physical toolkit

TESTGROEP A



TESTGROEP B



06 CONCEPT VALIDATION

This chapter describes the pilot of the toolkit as the final evaluation for the proposed concept. The chapter elaborates on the approach, gained insights, recommendations and limitations of the toolkit.

6.1 Evaluation pilot

This concept evaluation helps to answer the research question: How can AI experiments enable the formulation of policies for human-centered AI applications? The setup of the evaluation is explained in this chapter. The results are described and subsequently some future recommendations are suggested as are some limitations of this project pointed out.

Aim of the pilot (why)

As final activity, a pilot is held to test the toolkit with a representative experience panel of stakeholders who are affected by (and thus should be involved in) the process of formulating policies, guidelines and levers for responsible use of AI adoption. The aim of this pilot test was to validate if the toolkit elements are fulfilling the proposed design goals stated in Chapter 4.3: Toolkit Design goal.

Participants (Who)

The invite for the pilot was sent to:

- government representatives (the members of the AI team, innovation advisors and legal experts within the ministry) + an executing organizations representative (the Police AI lab)
- academic representatives (professors in the philosophy of technology and ethics field)
- business representatives (a Big Tech consultant, AI startup co-founders)
- society representatives (design students and voluntary citizens with little expertise in AI)

Set-up (How)

The entire toolkit validation consisted of three validation milestones as a part of the iterative process. The first round of validations consisted of a generative validation interview with stakeholders using concept visualisations to collect the first thoughts on the toolkit. The second round of validations included going through the canvasses with friends and relatives to observe the intuitive use and understand where more guidance is needed in the conversation flow. The final round of validations comprised experiencing the toolkit elements in a real life simulation. For an optimal pilot set-up, a fixation of the variables influencing the toolkit have been met by testing the toolkit in the chronological order, with the real (potential) users of the toolkit, in the expected timeframe of three weeks.

Time Frame (When)

The final concept validation was held in parallel with two test groups going through the same steps. As the toolkit requires multiple thought exercises, it is decided to split carrying out the complete toolkit into three validation rounds. Test Group A was scheduled for three repetitive sessions on Fridays from 10:30-12:00, and test group B in the afternoon from 14:30-16:00. In the days between the sessions, there was space to reflect on the learnings. At the start of a new session, a moment was provided to retrieve the insights from the previous session. The time between the sessions of Test Group A and B provided the possibility to iterate the canvases or set-up of the session based on the insights from group A, to immediately test with group B. The agenda for each session is shown in Figure 55.

USED AGENDA

Welkom! De agenda:

- 5 min: Introductie rondje
- 5 min: Beschrijving aanleiding Toolkit
- 10 min: Plan komende 3 sessies + uitleg Miro
- ~50 min: Doorlopen 2 canvassen + 5 min koffie break
- 10min: Discussie gebruik & Afronding

Figure 55: The agenda

Process (what)

The final pilot was held in the MIRO platform, so attendance of the various stakeholders could be easily accomplished. The agenda of the pilot consisted the following elements:

The session evaluation consisted of the following set-up (see Figure 56)

SESSION EVALUATION SET-UP

SESSIE EVALUATIE

PRODUCT

"De toolkit is wenselijk" (e.g. meerwaarde, aantrekkelijk, wow-effect) Toelichting:

"De toolkit is haalbaar" (e.g. in gebruik, structuur, systeem, tijd) Toelichting:

"De toolkit is levensvatbaar" (e.g. structureel, toekomst, draagvlak) Toelichting:

Wat zijn ideeën om de toolkit op deze bovenstaande punten hoger te laten scoren?

Overige opmerkingen

PROCES

Wat ging goed?

Wat kan er beter?



Figure 56: an visual overview of the pilot process and evaluation

6.2 Key insights

Feasibility

[CHECK] *The toolkit offers the ability to collaborate with a variety of stakeholders from remote (due to Covid pandemic), offering the potential of using the toolkit online*

Proof: The pilots have been completed via the online platform MIRO and the video conferencing program Webex, to which all participants had or gained access easily. The platform link is still available, all needed tooling is included, and the program offers the possibility to redo and undo changes, offering the possibility to endlessly continue using the online version of the toolkit.

[CHECK] *The toolkit offers the ability to empathise with stakeholders with a design artefact*

Proof: A participant about the Citizen perspective card: *“Very interesting to see it from a practical point of view, it really helps to look at the question from another angle.”* Other participants mentioned: *“Empathizing with other perspectives like this is a refreshing method”, “A great mirror of your own frame of reference and desired image”, “Empathy is crucial”, and “This way of empathizing with perspectives is a really valuable experience.”*

[CHECK] *The toolkit and needed software have a low entrance-barrier, meaning no prior knowledge or additional steps to require or purchase components.*

Proof: One participant mentioned finding the difference between ‘implementation’, ‘policy’, and ‘legislation’ a rather complicated question due to the abstractive terms. Also, the software in collaboration with Webex caused some computing power issues. However, with a short elaboration on framing these definitions, the canvasses were intuitively used. Moreover, the canvases have been adjusted in co-creation to increase its visual guidance. *“Engaging in a conversation is ‘easy’ practice. There are no expensive start-up costs, so definitely a low barrier to try.”* (participant)

Viability

[CHECK] *The toolkit guides a structural agile approach, using canvases to guide thought exercises and safeguard iteration in the process.*

Proof: Multiple participants mentioned that the structure was really helpful. *“I would like to test these canvases as well with my colleagues, as I believe this structure is really helpful.”* (pilot participant from the Police AI lab) Others mentioned: *“The MIRO board works very well, is refined and provides immediate feedback”, and “It can be used more often as every situation is different, so when thinking from a perspective you get a lot of insights”.*

[CHECK] *The toolkit incites to act, guiding the creation and testing of minimum viable policies for validation.*

Proof: Participants mentioned the feeling that they would like to start multiple research projects to understand which policy option would be the most successful. The group came up with realistic experiments to measure the effect (ranging from a poster to fake camera’s with a notification). *“Pre! It really encourages people to think broadly, outside of the box!”* (participant)

[MODERATE] *The toolkit supports the creation of Human-Centered AI using AI experiments as references to define policies.*

Proof: It supports the creation, but actual execution of the experiments are needed to monitor the success of the policy. It also requires a good moderator with knowledge on the use of the toolkit. Participants said: *“This is relevant in many different contexts, even broader than the context of AI”, and “There is a lot of demand for something so practical.”*

[CHECK] *The toolkit facilitates knowledge building my means of eliciting new information.*

Proof: Also proof of the ability to emphasize: *“I only see the effect now if that privacy activist does not feel heard. I could not have imagined that position of power before. This is new gained information for me now that we have drawn those connections.”* And another commented: *“This toolkit is especially an added value on awareness and stimulates realization”.*

Desirability

[CHECK] *The toolkit is inviting to use as the appearance is attractive.*

Proof: The pilot participants responded extremely positive. Multiple times participants mentioned the visualisations are very inviting. Some quotes from the participants: *“It is also very cool visually!”, “Very beautiful and clear design.”, and “It looks sexy, which is exactly what is needed for it to work”*

[CHECK] *The toolkit offers a fun experience, including an ‘invite for play’ element.*

Proof: The toolkit raised enthusiasm. Participants responded with: *“The canvases in combination with those cards generate a lot of discussion. Fun! I can talk for hours, so it’s good to set a time limit.”, and “No boring text, nice and interactive.”*

[CHECK] *The toolkit has the wow-factor, introducing a new way of working.*

Proof: *“I think this is really helpful. I would even like to invite more perspectives at the table, that would be even more valuable”, and “I would like to introduce and test this toolkit with my fellow colleagues at the Lab”.*

[MODERATE] *The toolkit provides a feeling of control, offering the ability to act fast and adaptively.*

Proof: For policymakers that are new with the software, a feeling of control was not yet experienced. This is expected to increase after a few guided experiences with the toolkit.

Other insights relevant to mention

Elaborate on the acronym CERTIFIED in the introduction to avoid the association with making something ISO certified. *“For people (outside the organisation) who are not familiar with the Certified approach this might be a bit confusing. My first association is a progress with the aim to get something ‘ISO certified’”* (Senior UX Designer, National Police)

Iterations have been made on the toolkit composition. Adding the visual ‘>’ cues increased the intuitive flow of the canvases. The text above in the canvases was not always read at first glance, therefore the text size is enlarged. Also, more civil servant jargon is added, and English words avoided to increase the comprehensibility and to relate the toolkit to their own daily practices.

Questions about ‘How to maintain innovation in a conservative environment?’ and ‘How would the added value of the toolkit remain if the creator is not present in the session itself?’ arose, emphasizing the need for a solid foundation for sustainable use.

The introduced ‘Minimum Viable Policy’ appeared still hard to grasp. Hence, during the MVP and the Test canvas, it was difficult to distinguish if the dialogue was about the policy or the AI experiment.

Filling out the canvases worked best for a group of 3-5 participants. The more participants, the more time delay was experienced. Hence, timekeeping is crucial and it is advised to add one facilitator to the session who can be solely held responsible for the timekeeping. Moreover, consider additional time to go through the insights of the previous canvases when the sessions are spread over multiple days or new participants attend the session.

An interesting quote by an interviewee worth to mention: *“When formulating an experiment, you actually want to include more research questions than the ones determined by the initiators. These questions could be some that are relevant to other stakeholders as well. You could also take policymakers along in the experiment. This is very beneficial for policymakers as it creates a close collaboration and these policymakers are involved two years earlier than they would otherwise be addressed. They will not immediately say ‘that is not allowed’, and by the time you have finished your experiment, then all involved parties have already learned. In that case, they have already thought about the policy and legislative frameworks that may be needed. So either make the policymaker an employee of the experiment or have the policy team provide a research question.”* - Interviewee

Enthusiasm was also registered during the various dialogues prior to the final pilot. Some quotes which emphasized the added value of this research:

- *“Would be great to make this toolkit part of the kickstart sessions or part of the (not yet existing) onboarding program”* (Program Manager Technology, Ministry of Justice & Security)
- *“Would be a great addition to the DPIA!”* (Quartermaster AI, Ministry of Infrastructure and Water management)
- *“It’s inviting, it looks very nice and inviting. I am very enthusiastic to work with it. Definitely like to work this structured”* (Senior Policy advisor, Ministry of Justice & Security)
- *“I definitely want to add those perspective cards to my toolkit list”* (Project Lead ‘Opgave Gericht werken’, Ministry of Justice & Security)
- *“Yes! Just write my name down. I fully support this structured way of working with this toolkit. I wish we would always work like this.”* (Jeroen Hak, agile expert, Ministry of Justice & Security)
- *“Cool! It looks great! There are many methods in it, so be careful it doesn’t get overwhelming. But I would like to receive a copy and stay informed!”* (Futures Researcher at Netherlands Study Centre for Technology Trends)

07 RECOMMENDATIONS & DISCUSSION

In this chapter, the final conclusions are made, and the recommendations for further implementation and research are given.

7.1 Discussion & Limitations

The initial research question stated at the beginning of this project was:
“How can AI experiments enable the formulation of policies for human-centered AI applications?”

which eventually lead to the formulation of the design challenge:
“Design a normative approach for agile policymaking, offering AI policymakers a tool to demystify the debate on human-centered AI”

This question came along with several sub questions that are answered during the explorative study that consisted of a literature review, generative sessions and field study. The answer to these questions led to the creation of the CERTIFIED manifesto and a toolkit, that is evaluated by the employees of the ministry and external parties.

Looking at the research question, it can be concluded that examples of AI experiments invite representatives of society to start the dialogue on demystifying the turning point on what is defined desirable by society. The manifesto provides the anchor for organizational change, where the toolkit is the first attempt to act according to the stated ‘rules of the game’. Emphasizing the metaphor provides a surrounding and mindset to embrace the uncertainty the journey of exploration onholds.

By following an human-centered design process, the deeper motivations of the people were discovered. By implementing the motivational elements of discovery, structure and transparency in the execution of the toolkit, an engaging experience was created. During the journey people could explore the act of proactive policymaking. However, the toolkit should first be evaluated over a longer period of time, in order to find if it can establish a structural part in the process of policy formulation.

This toolkit shows the ministry a possible and promising approach to act fast on regulating new policies. A limitation would be that it includes multiple elements. This might be overwhelming for an user at first glance. Providing one tool as an outcome would have increases the potential for implementation, as it would provide a clear focus and understanding of the tool. When offering a set of tools, it’s much harder to grasp the use of the tools in practice.

Spreading the manifesto acronym within the department can be a first step towards organizational change, which fits the vision to become an adaptive and resilient government. The first steps to gain awareness for the CERTIFIED manifesto have already been taken by including a wide group of stakeholders within this research. The proposed approach serves as an answer to questions currently stated in the strategy formulation for the ministry. This research provides a way to demstify public values and offers the practical tool to set things in motion. Hence, this research ties in with the urgency and provides an answer.



Figure 57: presenting a part of the toolkit during the conference: INNOvember

By publicly sharing an article to spark critical thinking, the first seeds have been planted in the need to create Minimum Viable Policies. Sharing the article throughout the organisations intranet and hosting webinars already gained traction. The card deck has already been added to the list of tools offered by the “Opgave Gericht Werken” community. I had the opportunity to take a stage with my research as a public speaker in a webinar of the INNOvember symposium (see Figure 57), resulting in requests for the card deck by various parties including the Royal Military Police. Enthusiasm has also been clearly announced by the director, as a result of which the manifesto and the toolkit are presented to the highest administrative body. This was prior to an invitation to join the panel discussions on formulating a new strategy for the ministry. This act indicates that enthusiasm is growing at the place where the needed mandate is established.

The Perspectives Carddeck proved to be a great method to gain consciousness about the potential consequences and concerns of other stakeholders. Policymakers experienced the blind spots in their thinking patterns and made them more aware of their own perspectives. During an evaluation session I was excited to hear that it triggered a policymaker to actually reach out to the real representative on his card, to ask and gather the information that appealed unknown to him.

A noted limitation is the security of the used software. The pandemic already showed its effect in the shift towards working from remote. We’ve become more tech-savvy, collaborating in online brainstorm tools and meetings. Therefore the online MIRO version offers great possibilities, however it is important to check the access to the board and the security. In particular as most discussions within the Ministry of Justice and Security consider potential sensitive information.

7.2 Recommendations

In this section, recommendations for future research are provided, on both practical (the design tools) and high-level (the transition for the government).

First of all, this thesis emphasis on the transition towards an adaptive government. As mentioned in Chapter 3, we cope with lots of uncertainties that all take place within the rapid pace of tech development. This calls for the need of an agile way of working. In particular, it calls for the implementation of the CERTIFIED approach. The toolkit introduces just one way to achieve this change, but there are many more opportunities to discover. The ministry should not lose sight of the longterm goals, swayed by the issues of today. The key message for using the toolkit is to gather various perspectives around the table for a cross-fertilization of knowledge on the desired future (regarding the use of certain AI systems). Not only well-known experts, but also the experience of the identified affected ones. Let curiosity guide you to the streets to validate your assumptions right after a session on 'how to retrieve public values' ended. Facilitate quick sprints to retrieve feedback as early as possible. I hope this document spreads throughout the organization by early-adopters, as a means to inspire you to become the pioneer that will guide others towards an adaptive and resilient government. It's time to change, and as the Covid-19 pandemic showed we widen our window of acceptability regarding policies in times of crisis, this is your call. Next, I present some ideas to further explore.

Turn policy teams into Multidisciplinair Intervention Teams

The 'Multidisciplinair Intervention team' already proved its value. A first step would be to increase a team's self-organizing capability (independance), by including an IT expert and Design expert in each policy team. Regarding the IT'er, this would not only help filling the gap to solve some feasibility challenges in organizational development right away, but also offers a new practical perspective to policymaking with expertise of the challenges in IT management. Regarding the Design expert, this would help communicating ideas and policy options with quickly generated and easy understandable visualisations. Policy povotypes or MVP test set-ups can be created in a few minutes. In addition, the Design Thinking approach to combat wicked problems has also proved its value from a holistic viewpoint. Design experts could share and guide these design methods to learn new flexible ways of working.

Granting time for the Long-term passion projects

Where most of our time currently goes to solving a slip of the tongue of a minister, politics shouldn't lose grip of the long term practices. As there is nobody who has an answer to the question 'Who owns the weather?', there should be the opportunity for vollunteers to take this responsibility. Meaning, provide structural space for proactive thinking. Start tomorrow with a 10% time allocation for future exploration. Imagine what a desired future would include and translate this into passion projects. This also requires an organizational structure that grants time for addressing questions that end up between shore and ship. A more deliberate approach would be to allocate 50% of the time for short-term goals, and reserve 50% for the long-term goals.

Launching a Governmental Creative Design Space

A compass as metaphor is the first step in providing a playful source of inspiration in the physical environment, but this can be taken to a larger scale. The effect on the organizational culture by creating an actual "governmental creative design space" or "Dutch Policy Lab"

would be a recommended topic to further explore. It could be a concept inspired by the UK Policylab or the Danish DesignLab. Already naming it a 'Lab' provides a space for experimentation with a culture that embraces failure. Researching this opportunity can lead to great insights for organisational learning. A normative collaboration with Design Thinking experts could both help shape this concept as well as help tackling the wicked problems.

Exploring a Freewheel law

Best case scenario, there should be a 'law for exploration', which should not be 100% open, but describes, for example, what on a micro-directive level at least should "not" happen in any case and provides a free-pass for a certain timeframe. The challenge, according to an interviewee, for such an experimental law is the legal construction. "*Test to Learn* is a legal minefield." It would be interesting to propose a vision with scenarios on how the government of the future would function with the introduction of *Freewheel laws*.

Creating policy provotypes

Avoid the theoretical discussions today by following the Guided Ethics framework, and starting to create provotypes. Albeit a photoshopped image with a provoking future scenario (e.g. of an AI system providing you tailored safety advice if you accept to give access to personal security cameras and being watched on your personal property). Each department could hire a graphic design intern to create these kind of provotypes. Let's give these potential future scenarios, representing value tensions, a central point of attention in the dialogue. And to collect even more insights on how an innovation 'lands' and can be used, so we should 'experience' it first with a Minimum Viable Policy test set-up to avoid the theoretical discussion and describe according to the experienced behaviour towards the prototype.

Measuring the value of creating Minimum Viable Policies

The published article already sheds light on the potential definition of a Minimum Viable Policy and the desire for it has been identified. Further research could be focused on how people perceive this term, what it could mean to them, and how such policy experiments could be prototyped. It could also be interesting to further improvement of the canvases leading to the MVP tests, or a study to measure the gained knowledge in comparison with current statistics.

Goal-oriented laws with micro-directives

Research on how goal-oriented laws in combination with tailored micro-directives would be interesting to further explore. How could these micro-directives be implemented and what it would mean for the government? If a law would only formulate a goal (i.e. "minimize road accidents") and context factors are collaboratively defined, a program could run a set of micro-directives. If context factors change, the micro-directive is adjusted, which is an iterative process, while the fundamental basics remain: realizing fewer accidents on the road. These micro-directives are concrete and can be iteratively tested quickly, minimizing the gap of abstraction to concrete practices.

08 PERSONAL REFLECTION

In this chapter, I reflect on my personal ambitions during my graduation journey and on the thesis results.

8.1 Personal ambitions

My personal motivation for this research was to seize the opportunity to delve into research on AI technology and its related challenges. With my passion and endless curiosity, I wanted to extend my multidisciplinary skill set to new areas which I haven't explored yet. If I would have chosen to graduate in a familiar environment of design agencies with like-minded people, I wouldn't have learned how to explain my role as designer and convince how I could be of added value. As there was no previous collaboration with the IDE faculty, I wanted to explore this collaboration and experience how the role of a designer was perceived. I wanted to create a fruitful collaboration with mutual learning. I would be able to learn about the political context and the organizational complexity of large public organizations which would help me decide on my future career, whereas I could show the power of Design Thinking in a text-driven organization. I have been granted a unique opportunity to create impact on a large scale for a societal challenge regarding my role in the AI team.

Design competences I wanted to express and improve considered:

- *being able to apply design methods in research*, which I think I covered extensively.
- *making insights representable and a supporting tool for collaboration*, which I managed quite well considering the positive feedback on the visualisations.
- *apply creative facilitation skills in practice*, which turned out to be mostly online in Miro. Regarding the final validations, I could have spent more attention to timing.
- *dare to choose methods, iterate quickly and embrace feedback*. As I had trouble distinguishing main from subsidiary issues, choosing was tough sometimes. In addition, I have been my own critic which provided a high threshold to present unfinished work.
- *create alignment in communication with involved stakeholders, and maintain true to own design beliefs*. I involved many stakeholders in my research, resulting in a great network and 'ownership' for my research and the implementation. It has been a strategic choice to generate this 'ownership' throughout the proces. I truly believe as designer it is crucial to co-develop a deliverable. This will increase its value and its potential for being implemented after my research has finished. Aside from showing the value of using a human-centered approach myself, I have created an opportunity for others to experience this approach themselves.

My points of attention to grow as a person considered:

- Don't let ambition enable own unrealistic expectations and planning
- Find peace in the unknown and intangible process by trusting my intuition.
- Find confidence in my own decisions without continuous confirmation from others
- Preserve a holistic perspective on the thesis, don't get too distracted by detail.
- Not only the outcome weighs in pride, but the road towards it weighs equally.

During this journey, I experienced some trouble finding the balance between work and wellbeing. Along the way, I have grown massively on this part by indicating boundaries in time and choosing to skip the habit of structural overworking. It has been a rollercoaster, but I wouldn't have it otherwise. This period of time has taught me a great amount of valuable lessons, for which I'm grateful I've learned them in this phase of my life.

8.2 Thesis result

How cliché it might be, one of the biggest ambitions was that the results of this thesis would not be left on the 'pile to be forgotten'. As for the manifesto, change only happens when the need, the mandate and the resources come together. To do so, I included many experts in my research and tried to get as much feedback as possible early in the process. This way, the feeling of ownership would be created by the included parties. Actively collaborating and updating the diverse parties resulted in the demand and the mandate to bring the toolkit to the next level and lower the barrier for implementation.

This project includes many insights on the complexity of the tasks of policymakers. One of the most valuable shifts I experienced reflecting on the start and end of this project, was that from this moment on, problem statements include a person instead of 'society' as a whole. The question: *'But for whom is this a problem, and why does this representative actually experience it as a problem?'* was like music to my ears. That point made me realize that taking my colleagues along in the human-centered design process established a major effect. Also, the desire to further incorporate Design Thinking methods is acknowledged. Finally, my contributions as a designer within the team has not gone unnoticed. The desire for a successor provides a solid ground that the added value is of great extent.

To conclude, throughout the journey I felt the enthusiasm and willingness of the involved parties to further explore the concept of agile policymaking. It resulted in attending the INNOvember symposium where I took the stage to share the insights of my research, being invited as representative of the AI team to a panel of Centrale Eenheid Strategy to advise and share new ways to formulate strategies. I have been asked to give inspiring presentations at the police, and validation participants mentioned the desire to test the toolkit with their colleagues from other departments (e.g. Datalab, Politie AI lab). The article about the Minimum Viable Policy is being widely spread throughout the in- and external channels of the police, JenV Connect and LinkedIn. To conclude, the MIRO platform that I introduced at the start of this research, is now used on a daily base as online collaboration tool, to collect insights from meetings, facilitate brainstorm sessions, and to stay up-to-date about the teams daily priorities. This could not have happened if I hadn't had the opportunity to be this closely involved with the AI team. The team has been a solid ground (especially in this COVID-pandemic times) to connect with relevant people, exchange knowledge, and make the impact I wished to achieve. I foresee great expectations of the AI team in the near future.

REFERENCES

Visuals in this thesis have been created from resources of Freepik.com. Below the sources used in this thesis have been listed:

A

AI Impacts (2015) Predictions of Human-Level AI. Retrieved from: <https://aiimpacts.org/predictions-of-human-level-ai-timelines/>

Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*. <https://doi.org/10.1177/2053951720949566>

B

Biber, E., Light, S. E., Ruhl, J. B., & Salzman, J. (2017). Regulating business innovation as policy disruption: From the model T to Airbnb. *Vand. L. Rev.*, 70, 1561.

Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M. & Kern, J. (2001). The agile manifesto. Retrieved from: <http://agilemanifesto.org>

Beruti, F, Nel, P & Whiteman, R. (2020). McKinsey report. Retrieved from: <https://www.mckinsey.com/business-functions/operations/our-insights/an-executive-primer-on-artificial-general-intelligence#>

Boin, A, P 't Hart, E Stern and B Sundelius (2005) *The Politics of Crisis Management*. Cambridge: Cambridge University Press.

Bohlen, J.M. & Beal, G.M. (May 1957). "The Diffusion Process" (PDF). *Special Report No. 18*. 1: 56-77.

C

Cath C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 376(2133), 20180080. Retrieved from: <https://doi.org/10.1098/rsta.2018.0080>

Calabretta, G., Gemser, G., & Karpen, I. (2016). *Strategic Design: Eight essential practices every strategic designer must master*. Bis publishers.

Carayannis, E.G. & Campbell, D.F.J. (2009). "Mode 3' and 'Quadruple Helix': toward a 21st century fractal innovation ecosystem". *International Journal of Technology Management*. 46 (3/4): 201. doi:10.1504/IJTM.2009.023374. ISSN 0267-5730.

D

Chui, M. & Kamalnath, V. (2020) An executive's guide to AI. McKinsey review. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/an-executives-guide-to-ai>

CNBC (2020) News article. Clearview AI Facial Recognition Europe. Retrieved from: <https://www.cnn.com/2020/06/11/clearview-ai-facial-recognition-europe.html>

Conklin, Jeff; Basadur, Min; VanPatter, GK (2007a). "Rethinking Wicked Problems: Unpacking Paradigms, Bridging Universes (Part 1 of 2)". *NextDesign Leadership Institute Journal*.

Conklin, Jeffrey (2006). *Dialogue mapping: building shared understanding of wicked problems*. Chichester, England: Wiley Publishing. ISBN 978-0-470-01768-5.

De Bono, E. (2017). *Six thinking hats*. Penguin uk.

De Nationale AI cursus. <https://app.ai-cursus.nl/home>

Deloitte. (2019). The rise of data and AI ethics. Deloitte Insights. Retrieved from: <https://www2.deloitte.com/us/en/insights/industry/public-sector/government-trends/2020/government-data-ai-ethics.html>

Deloitte. (2020). Deloitte Insights. Report on Governmental Trends. Retrieved from: <https://www2.deloitte.com/us/en/insights/industry/public-sector/government-trends/2020/government-data-ai-ethics.html>

Digitale Overheid (2020) Hoe denken Nederlanders over Artificiële Intelligentie. Retrieved from: <https://www.digitaleoverheid.nl/achtergrondartikelen/hoedenken-nederlanders-over-artificiele-intelligentie/>

DOMO (2020). Data never sleeps. Retrieved from: <https://web-assets.domo.com/blog/wp-content/uploads/2020/08/20-data-never-sleeps-8-final-01-Resize.jpg>

REFERENCES

Dorst, C. H. (2015). Frame innovation: create new thinking by design. *Design thinking, design theory*.

E

EU Commission (2020). "On Artificial Intelligence - A European approach to excellence and trust". EC Communications 65. URL: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

F

Ford, K. M., Hayes, P. J., Glymour, C., & Allen, J. (2015). Cognitive orthoses: toward human-centered AI. *AI Magazine*, 36(4), 5-8.
Frissen, V., Van Eck, M. & Drouwen, T. (2020). Hooghiemstra & Partners. Research Report on Supervising governmental use of algorithms. Retrieved from: <https://hooghiemstra-en-partners.nl/wp-content/uploads/2020/01/Hooghiemstra-Partners-rapport-Supervising-Governmental-Use-of-Algos.pdf>

G

Gunderson, L.H & C.S. Holling (2002) *Panarchy, understanding transformation in human and natural systems*. Washington, DC: Island Press, p. 15-21.

Halse, J. & Boffi, L. (2016). "Design interventions as a form of inquiry." In *Design Anthropological Futures*, edited by Rachel Charlotte Smith, Kasper Tang Vangkilde, Mette Gislev Kjærsgaard, Ton Otto, Joachim Halse and Thomas Binder. Bloomsbury.

H

Hao, K. (2018) 'Is it using AI?' The Algorithm, MIT Technology Review. Retrieved from: (<https://www.technologyreview.com/s/612404/is-this-ai-we-drew-you-a-flowchart-to-work-it-out/>)

Hasche, N., Höglund, L. & Linton, G. (2020) Quadruple helix as a network of relationships: creating value within a Swedish regional innovation system, *Journal of Small Business & Entrepreneurship*, 32:6, 523-544, DOI: 10.1080/08276331.2019.1643134

Head, B. W. (2008). Wicked problems in public policy. *Public policy*, 3(2), 101.

Hill (2020). *New York Times*. Retrieved from: <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>

K

Holierhoek, S. E., & Price, R. A. (2019, November). The Role of Design in policymaking. In *Conference Proceedings of the Academy for Design Innovation Management (Vol. 2, No. 1, pp. 999-1012)*.

Kalluri, P. (2020). Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature*, 583(7815), 169-169.

KCWJ. Rijksoverheid. Factsheet Beleidsregels. Retrieved from: https://www.kcwj.nl/sites/default/files/Factsheet_Beleidsregels.pdf

Kolk, A. (2019) Prototyping and Design Hacks. Slideshare. Retrieved from: <https://www.slideshare.net/InnovationAcademy/eia2019hk-prototyping-and-design-hacks-alar-kolk>

Kunz, Werner; Horst W. J. Rittel (1970). *Issues as Elements of Information Systems*. Heidelberg, Germany. CiteSeerX 10.1.1.134.1741.
Leeuw (2019). Big Data, AI en de wereld van Justitie en Veiligheid: Kansen en Uitdagingen voor Documentatie, wetgeving, de strafrechtketen en evaluatie-onderzoek, 2019

M

Middendorp (2019) Initiatiefnota Menselijke grip op algoritmen. Retrieved from: <https://janmiddendorp.nl/wp-content/uploads/2019/05/35212-2-Initiatiefnota-van-het-VVD-lid-Middendorp-Menselijke-grip-op-algoritmen-PDF.pdf>

Mommers. J & Rovers. E (2020). De Correspondent. Meer dan een inspraakavond, beter dan een referendum. Retrieved from: <https://decorrespondent.nl/11707/meer-dan-een-inspraakavond-beter-dan-een-referendum-met-deze-democratische-vernieuwing-kan-de-burger-zelf-beleid-maken/1290216763-753ad347>

Mul, S. & Werkhorst, K. (2020) De Digitale Transformatie vraagt om een adaptieve wetgever. Retrieved from: <https://esb.nu/esb/20059107/de-digitale-transformatie-vraagt-om-een-adaptieve-wetgever>

Musk, E. (2017) *The Verge*. Retrieved from: <https://www.theverge.com/2017/7/17/15980954/elon-musk-ai-regulation-existential-threat>

N

NOS (2020) News article. Hoe algoritmes de stad besturen. Retrieved from: <https://nos.nl/nieuwsuur/artikel/2355542-hoe-algoritmes-de-stad-besturen.html>

REFERENCES

P

Parool. (2020) Overheid stopt met gebruik SyRI na uitspraak rechter. Retrieved from: <https://www.parool.nl/cs-bbf3993a>

R

Rijksoverheid (2020) Chamber letter. TK uitwerking for breed offensief tegen georganiseerde ondermijnende criminaliteit. Retrieved from: <https://www.rijksoverheid.nl/documenten/kamerstukken/2020/06/18/tk-uitwerking-breed-offensief-tegen-georganiseerde-onder-mijnende-criminaliteit>

Rijksoverheid. Rijks Innovatie Community. Het ministerie van Justitie en Veiligheid ontwikkelt de samenwerking met startups door: het nieuwe normaal! Retrieved from: <https://www.rijksinnovatiecommunity.nl/actueel/nieuws/2020/01/30/startup-hub>

Rijksoverheid. Webpage Rijksoverheid.nl Retrieved from: <https://www.rijksoverheid.nl/ministeries/ministerie-van-justitie-en-veiligheid>
Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, 4(2), 155-169.
Rittel, H. W. (1988). The reasoning of designers. *IGP*.

Rosman. (2019) Wie zijn de onderzoekers van Bellingcat. Retrieved from: <https://www.ad.nl/buitenland/wie-zijn-de-onderzoekers-van-bellingcat-aa41ed00/?referrer=https%3A%2F%2Fwww.google.com%2F>

Rozenblum, R., Rodriguez-Monguio, R., Volk, L. A., Forsythe, K. J., Myers, S., McGurrian, M. & Seoane-Vazquez, E. (2020). Using a machine learning system to identify and prevent medication prescribing errors: a clinical and cost analysis evaluation. *The Joint Commission Journal on Quality and Patient Safety*, 46(1), 3-10.

RVO. Hoe werkt SBIR. Visited on July 6th 2020. Retrieved from: <https://www.rvo.nl/subsidie-en-financieringswijzer/sbir/hoe-werkt-sbir>

S

Sanders, E. B.-N. & Stappers, P. J. (2012). Convivial toolbox: Generative research for the front end of design. Amsterdam: BIS.

SAPAI (2019) Strategisch actieplan voor artificiële intelligentie. Beleidsnota, 8 oktober. Retrieved from: www.rijksoverheid.nl.

Satell, G. & Abdel-Magied. Y (2020) *Harvard Business Review*. AI Fairness Isn't Just an Ethical Issue. Retrieved from: https://hbr.org/2020/10/ai-fairness-isnt-just-an-ethical-issue?utm_source=linkedin&utm_medium=social&utm_campaign=hbr

Snowden, D. J., & Boone, M. E. (2007). A leader's framework for decision making. *Harvard business review*, 85(11), 68.

Staatscourant. (2019) Organisatiebesluit WODC. Retrieved from: <https://www.wodc.nl/organisatie/>

Sternberg R.J. The Theory of Successful Intelligence. *Review of General Psychology*. 1999;3(4):292-316. doi:10.1037/1089-2680.3.4.292

Stolterman, E. (2008). "The Nature of Design Practice and Implications for Interaction Design Research". *International Journal of Design*. 2: 55-65.

Szalek, B. Z. (2013). Some praxiological reflections on the so called 'overton window of political possibilities', 'framing' and related problems. *Reality of Politics. Estimates-Comments-Forecasts*, (4), 237-257

T

Tallacchini, M. (2009). Governing by values. *EU ethics: soft tool, hard effects*. *Minerva*, 47(3), 281.

TU Delft (2020). Participatory Value Evaluation (PVE). Retrieved from: <https://www.tudelft.nl/en/tpm/pve/>

U

UK Policy Lab (2014). Introduction to the UK Policy Lab. Slideshare. Retrieved from: https://www.slideshare.net/Openpolicymaking/policy-lab-slide-share-introduction-final?qid=adcbf9c0-ba85-4ad2-a99e-60a144b8d25d&v=&b=&from_search=1

V

Van den Hoven, J. (2020). Data Ethics Decision Aid (DEDA): A dialogical framework for ethical inquiry of AI and data projects in the Netherlands (manuscript)

Van der Steen, M., Van Delden, M., & Van Schaik, E. (2020) De opgave aan tafel. Retrieved from: <https://www.nsob.nl/denktank/overzicht-van-publicaties/de-opgave-aan-tafel>

REFERENCES

Van Leeuwen, M. (2019). Kwartiermakers plan, Article 3.2 AI, report of the board of directors 29-11-2019.

Van Schie, Gerwin; Westra, Irene; Schäfer, M.T. (2017) Schäfer, Mirko Tobias, van Es, Karin (eds.), The Datafied Society, pp. 183 - 200

Verbeek, P. P., & Tijink, D. (2019). Aanpak begeleidingsethiek: een dialoog over technologie met handelingsperspectief.

Verhage (2020). Implementing Ethical Artificial Intelligence: Approaches to Ethics-Based Governance. Retrieved from: <https://nlintheusa.com/wp-content/uploads/2020/10/Implementing-Ethical-AI.pdf>

APPENDICES

Appendix A: Overview of events for data collection

Appendix B: Extensive research on mimicking human intelligence

Appendix C: Context cards using in session to explore the value of AI

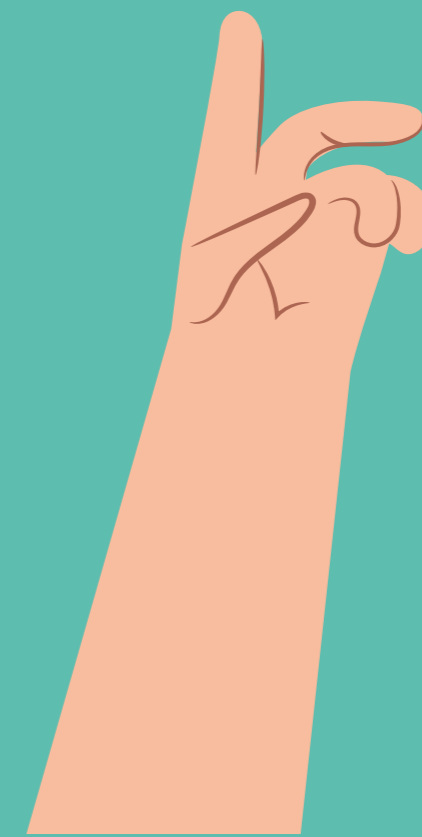
Appendix D: Dutch version of CERTIFIED Manifesto

Appendix E: Perspective Card deck cards

Appendix F: Toolkit Canvases

Appendix G: Trigger cards

*Ready to become a
CERTIFIED trainer?
Have a look at the tools
in the appendices!*



List of interviewees and participants is anonymized for publication and available upon request.

To learn

Learning is an overarching component in this case. There are a number of different forms of learning as applied to artificial intelligence. We can distinguish four purpose areas of application: description,

Descriptive analysis doesn't necessarily include AI as there is no learning component in the analysis itself. When a prediction or prescription element is present, there is a decision element involved. In this case machine learning occurs. Machine learning can be distinguished in reinforcement learning, unsupervised learning and supervised learning. (Rozenblum et al., 2020)

Reinforcement learning

An algorithm learns to perform a task simply by trying to maximize rewards it receives for its actions (eg. maximizes points it receives for increasing returns of an investment portfolio.

When to use it:

You do not have a lot of training data; you cannot clearly define the ideal end state; or the only way to learn about the environment is to interact with it.

How it works:

The algorithm takes an action on the environment (eg, makes a trade in a financial portfolio) It receives a reward if the action brings the machine a step closer to maximizing the total rewards available (eg, the highest total return on the portfolio)

The algorithm optimizes for the best series of actions by revising itself over time

Unsupervised learning

An algorithm explores input data without being given an explicit output variable (eg, explores customer demographic data to identify patterns)

When to use it:

You do not know how to classify the data, and you want the algorithm to find patterns and classify the data for you.

How it works:

The algorithm receives unlabeled data (eg, a set of data describing customer journeys on a website)

It infers a structure from the data

The algorithm identifies groups of data that exhibit similar behavior (eg, forms clusters of customers that exhibit similar buying behaviors)

Supervised learning

An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output (eg. How the inputs 'time of year' and 'interest rates' predict housing prices)

When to use it:

You know how to classify the input data and the type of behavior you want to predict, but you need the algorithm to calculate it for you on new data

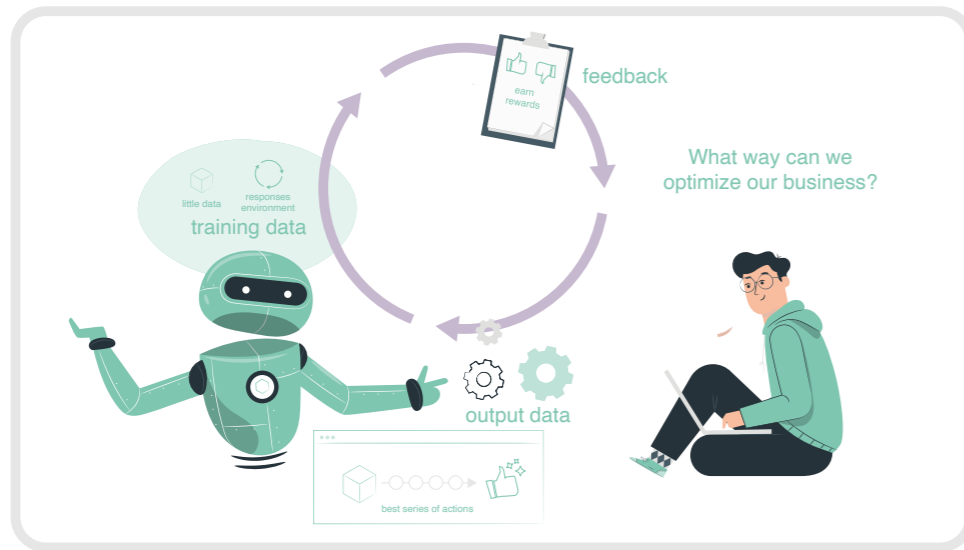
How it works:

A human labels the input data (eg, in the case of predicting housing prices, labels the input data as 'time of year', 'interest rates', etc) and defines the output variable (eg, housing prices)

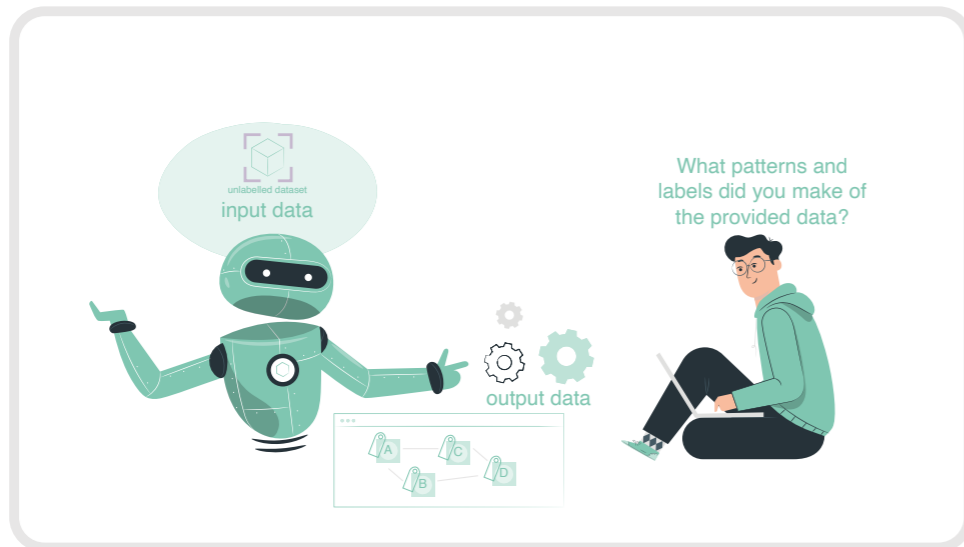
The algorithm is training on the data to find the connection between the input variables and the output. Once training is complete - typically when the algorithm is sufficiently accurate - the algorithm is applied to new data.

Deep learning

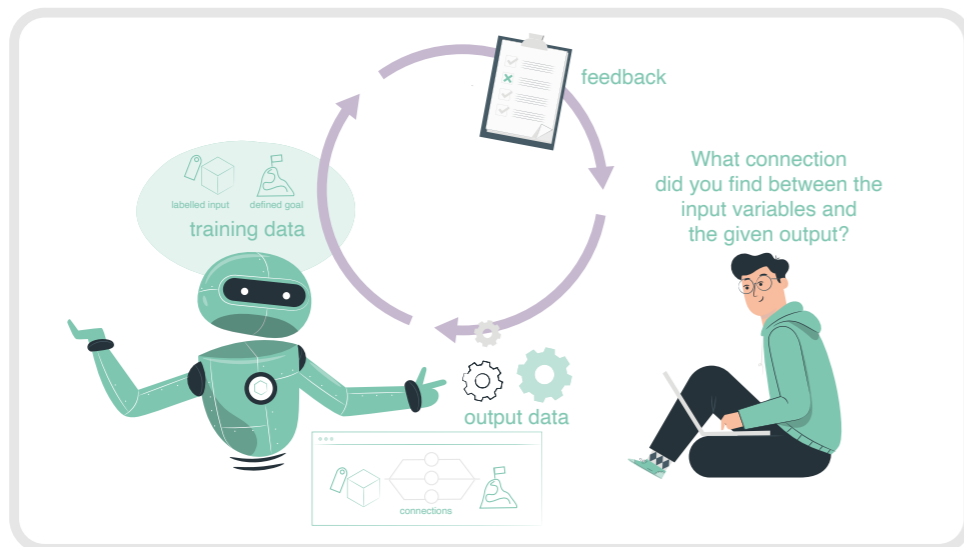
Deep learning is a type of machine learning that can process a wider range of data resources, requires less data preprocessing by humans, and can often produce more accurate results than traditional machine-learning approaches (although it requires a larger amount of data to do so). In deep learning, interconnected layers of software-based calculators known as "neurons" form a neural network. The network can ingest vast amounts of input data and process them through multiple layers that learn increasingly complex features of the data at each layer. The network can then make a determination about the data, learn if its determination is correct, and use what it has learned to make determinations about new data. For example, once it learns what an object looks like, it can recognize the object in a new image.



REINFORCEMENT LEARNING



UNSUPERVISED LEARNING



SUPERVISED LEARNING

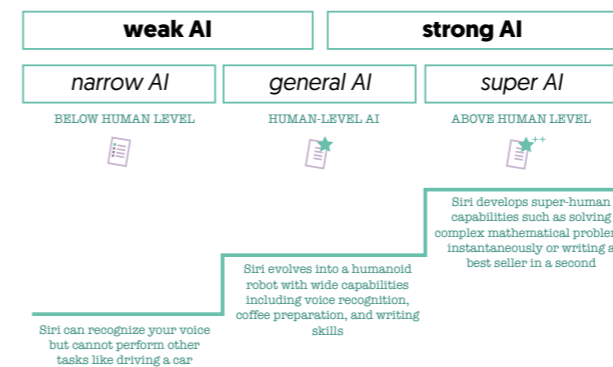
To reason

To reason is to draw inferences appropriate to the situation. Inferences are classified as either deductive or inductive. Deductive inferences guarantee the truth of the conclusion, which is common in mathematics. An example: "X must be either A or B. He is not in A, therefore he is in B." Inductive inferences guarantee truth of the premise that supports conclusion (without absolute assurance), which is common in science. An example: "Similar accidents of this sort were caused by A. Therefore this accident was caused by A."

To problem-solve

Problem solving by AI may be characterized as a systematic search through a range of possible actions in order to reach some predefined goal or solution. We can distinguish weak AI, by the special-purpose, from strong AI, which has a general purpose. A special-purpose method is tailor-made for a particular problem and often exploits very specific features of the situation in which the problem is embedded. A general-purpose method is applicable to a wide variety of problems. One general-purpose technique used in AI is means-end analysis—a step-by-step, or incremental, reduction of the difference between the current state and the final goal.

Instead of referring to weak or strong AI, the maturity level to problem solve can also be used to describe the relation to human intelligence:



Narrow AI: An AI system can do a specific task very well, such as making coffee, but if you're out of coffee beans, it can't ask the neighbour to borrow their beans.

General AI: An AI system evolves in a humanoid bot that uses AI in several areas from preparing coffee, to voice recognition to writing skills.


Super AI: An AI system outperforms humans on all levels

To perceive

Perception of the environment is closely related to recognition by our senses. The environment is scanned by means of various sensory organs and the scene is decomposed into separate objects in various spatial relationships. Analysis is complicated by the fact that an object may appear different depending on the angle from which it is viewed, the direction and intensity of illumination in the scene, and how much the object contrasts with the surrounding field. As humans have senses, an AI distracts the data from sensors. Sternberg (1999) stresses the difference between internal intelligence, the mind, and external intelligence, the environment. Intelligent behavior is determined in large part by the environment in which one lives, and the kinds of situations that this environment tends to present. This can be translated to the ability of the AI to connect to the sensors present in the environment. Where eyes represent vision, a camera can be combined with image processing or facial recognition. Computer vision can also analyse patterns to read text and respond with chatbots. Where ears represent sound, recordings can be combined with voice recognition software or Natural Language Processing to help transcribing. Where fingers represent touch, geo-location or movement sensors can be combined with smart robots to automatically pick up and move fragile objects. Where a tongue represents taste, heat sensors can detect and monitor a desired temperature. Note: in this research, there haven't been applications identified with a particular focus on taste itself yet. Where a nose represents scent, an electronic nose can detect chemical substances based on the volume of the distribution of gases present.



To process

Processing information is a way to interpret a language. A language is a system of signs having meaning by convention. In this sense, language need not be confined to the spoken word. Traffic signs, for example, form a mini language, it being a matter of convention that  means “hazard ahead” in some countries. There are three levels of semiotics to consider: Syntactics, Semantics, and Pragmatics. Syntactics deals with the signs, and their relation to other signs. This establishes a set of rules for combining signs, or a calculus. Syntactics is abstracted from the situation surrounding the information process, whether measurement or another information process. Semantics deals with the relation of signs to objects, and abstracts from specific situations. In effect it studies the communication, or information, process as a standard situation, with a standard set of conditions, no external influences, and a standard receiver, or user. Pragmatics includes all the specific factors, including personal and psychological, which make each communication/information event unique. This level accounts for the role of the user, the external influences, and the specific situation surrounding the information process.

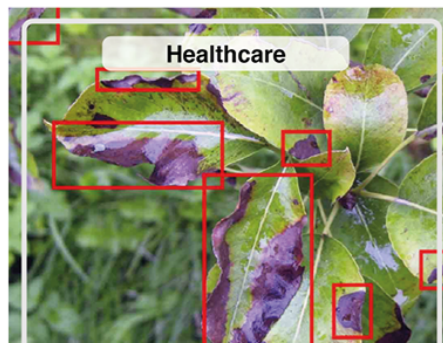
In general, we distinguish four areas of processing data by algorithms:

1. Descriptive - Analysis of “What is happening”: the algorithm provides a representation of what is being observed;
2. Diagnostic - Analysis of “Why is it happening” - the algorithm provides a probability or analysis of why something is occurring caused by positive, negative or predictive values, or calculates a particular probability used in work processes and / or decision making;
3. Predictive - Analysis of “What will happen” - the algorithm makes a prediction of what may happen or a probability / probability of a future action or event;
4. Prescriptive - Analysis of “What to do”: The algorithm determines and / or dictates the decision / action or execution.

Ingredients for intelligence

Humans have a set of morals that drive our codes of conduct. We drink, eat and sleep to provide our brains with the energy in order to think clearly. Our senses send information via stimuli to our brain. These aspects form the ingredients for our intelligence, which can also be translated to the case of AI. The three ingredients for AI are: data, algorithms and computing power. Data is generated by sensors. Algorithms are the rules the AI follows, created by developers based upon a certain input data and (un)defined output. Computing power is how fast a machine can perform an operation. Everything a computer does breaks down into math. A computer’s processor interprets any command you execute as a series of math problems. Faster processors can handle more calculations per second than slower ones, and they’re also better at handling really tough calculations. Computing power is generally estimated in terms of accuracy, efficiency and speed of executing computer program instructions.

 <p>Healthcare</p> <p>“An intelligent symptom checker.”</p> <p>Description: Buoy Health is an AI-based symptom and cure checker that uses algorithms to diagnose and treat illness. Here’s how it works: a chatbot listens to a patient’s symptoms and health concerns, then guides that patient to the correct care based on its diagnosis.</p> <p>Source:</p>	 <p>Healthcare</p> <p>“Better insights in patient’s real-time needs.”</p> <p>Description: Enlitic develops deep learning medical tools to streamline radiology diagnoses. The company’s deep learning platform analyzes unstructured medical data (radiology images, blood tests, EKGs, genomics, patient medical history) to give doctors better insight into a patient’s real-time needs.</p> <p>Source:</p>	 <p>Healthcare</p> <p>“Earlier cancer detection with AI.”</p> <p>Description: Freemove uses AI in screenings, diagnostic tests and blood work to test for cancer. By deploying AI at general screenings, Freemove aims to detect cancer in its earliest stages and subsequently develop new treatments.</p> <p>Source:</p>
 <p>Business</p> <p>“Predicting stock market changes.”</p> <p>Description: Banks use artificial intelligence systems today to organize operations, maintain book-keeping, invest in stocks, and manage properties. AI can react to changes overnight or when business is not taking place. Kensho is an intelligent computer system widely used by stock traders and investors to automatically analyse portfolio performance and predict market changes.</p> <p>Source:</p>	 <p>Entertainment</p> <p>“AI assistant for life comfort.”</p> <p>Description: Amazon is using neural networks and natural language models to improve Alexa’s speaking style, give it new voices, and even notice that a user is unhappy with it.</p> <p>Source:</p>	 <p>Business</p> <p>“Facial recognition to create large database of people.”</p> <p>Description: The company Clearview AI uses ML software to compare faces on camera images to public photos and videos on the internet. Clearview has scraped social media and created a database of three billion photos with associated names. This creates data monopoly and risks for hackers.</p> <p>Source: Volkskrant, 6 mrt 2020</p>
 <p>Healthcare</p> <p>“Identify and predict harmful bacteria with accuracy.”</p> <p>Description: Beth Israel Deaconess Medical Center are using AI-enhanced microscopes to scan for harmful bacterias in blood samples at a faster rate than is possible using manual scanning. The machine learned to identify and predict harmful bacteria in blood with 95% accuracy, offering the opportunity to diagnose potentially deadly blood diseases at a very early stage.</p> <p>Source:</p>	 <p>Healthcare</p> <p>“AI-enabled assistant for various detections.”</p> <p>Description: Zebra Medical Vision provides radiologists with an AI-enabled assistant that receives imaging scans and automatically analyzes them for various clinical findings it has studied. The findings are passed onto radiologists, who take the assistant’s reports into consideration when making a diagnosis.</p> <p>Source:</p>	 <p>Healthcare</p> <p>“AI to identify and develop new medicines.”</p> <p>Description: BioXcel Therapeutics uses AI to identify and develop new medicines in the fields of immuno-oncology and neuroscience. Additionally, the company’s AI drug re-innovation program employs AI to find new applications for existing drugs or to identify new patients.</p> <p>Source:</p>



Healthcare

“Treating rare diseases with AI.”

Description: BERG is a clinical-stage, AI-based biotech platform that maps diseases to accelerate the discovery and development of breakthrough medicines. By combining its “Interrogative Biology” approach with traditional R&D, BERG can develop more robust product candidates that fight rare diseases.

Source:

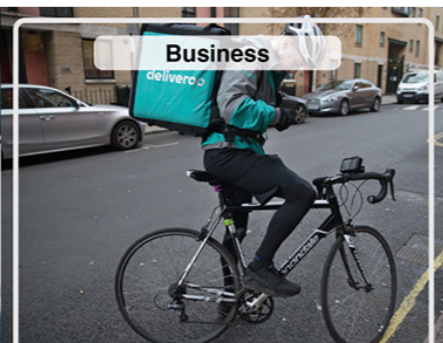


Research

“Find useful source within research papers.”

Description: Robot arm performs better

Source: <https://techcrunch.com/2019/03/08/polyai/>



Business

“Deliveroo optimizes profit by algorithm decision-making.”

Description: Mya Systems has developed a digital assistant that can streamline the recruitment process in various ways, view resumes, gather candidate information, and schedule job interviews. Previously, recruiters turned out to be overwhelmed by “mindless” work. They believe that these chatbots help with the bureaucratic and repetitive aspects of the recruiting process.

Source: <https://techcrunch.com/2019/03/08/polyai/>



Legislation

“Algorithms identify repeat offenders better than judges.”

Description: A new study in Science Advances more rigorously confirms that algorithmic judgements may in fact be more accurate than humans. Of concern, though, is that given the stakes involved — future crimes, a defendant’s freedom or continued incarceration — they’re still not reliable enough to ensure that justice is truly done and that tragic mistakes can be avoided.

Source: <https://advances.sciencemag.org/content/6/7/eaaz0652/tab-pdf>



Entertainment

“Generating deepfakes for creativity.”

Description: Music from Michael Jackson, artworks inspired by collections from Van Gogh, use a randomizer to generate voices, artworks, pictures. Based upon the input data patterns is recognized and transformed into a deepfake, which can be both positively used (i.e. historical voice for optimal museum experience) and negatively used.

Source: <https://news.artnet.com/market/first-ever-artificial-intelligence-p-677eaaz0652/tab-pdf>



Healthcare

“Cloud based digital drug discovery.”

Description: Combining AI, the cloud and quantum physics, XtalPi’s ID4 platform predicts the chemical and pharmaceutical properties of small-molecule candidates for drug design and development. Additionally, the company claims its crystal structure prediction technology predicts complex molecular systems within days rather than weeks or months.

Source:



Healthcare

“Increasing access to healthcare.”

Description: Babylon uses AI to provide personalized and interactive healthcare, including anytime face-to-face appointments with doctors. The company’s AI-powered chatbot streamlines the review of a patient’s symptoms, then recommends either a virtual check-in or a face-to-face visit with a healthcare professional.

Source:



Healthcare

“Real-time patient flow optimization.”

Description: Qventus is an AI-based software platform that solves operational challenges, including those related to emergency rooms and patient safety. The company’s automated platform prioritizes patient illness/injury, tracks hospital waiting times and can even chart the fastest ambulance routes.

Source:

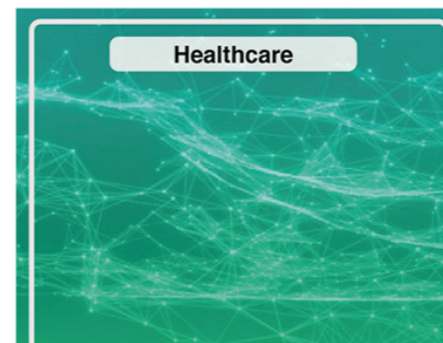


Business

“Self-driving cars with surrounding detection.”

Description: Advancements in AI contribute to the creation and evolution of self-driving vehicles. As of 2016, there are over 30 companies utilizing AI into the creation of self-driving cars (i.e. Tesla, Google, Apple). Google has been working on an algorithm with the purpose of eliminating the need for pre-programmed maps and instead, creating a device that would be able to adjust to a variety of new surroundings.

Source: <https://techcrunch.com/2019/03/08/polyai/>

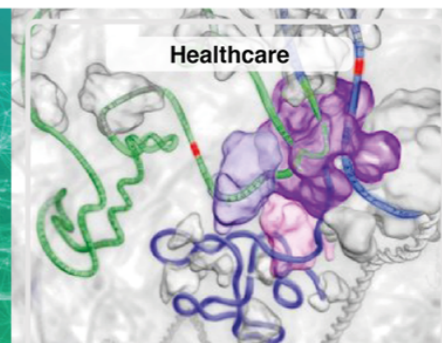


Healthcare

“Neural network for clinical trials.”

Description: Atomwise uses AI to tackle some of today’s most serious diseases (i.e. Ebola) whose neural network, AtomNet, helps predict bioactivity and identify patient characteristics for clinical trials. It screens between 10 and 20 million genetic compounds each day and can reportedly deliver results 100 times faster than traditional pharmaceutical companies.

Source:



Healthcare

“Unlocking therapies by AI feedback.”

Description: Deep Genomics’ “Project Saturn” analyzes over 69 billion different cell compounds and provides researchers with feedback. Their platform includes a library of 1000 compounds that are experimentally verified to manipulate cell biology as intended. It’s a toolkit for controlling cell biology along crucial pathways, rapidly unlocking therapies with greater potential.

Source:



Healthcare

“AI-as-a-service: Streamline patient’s experience”

Description: Olive’s AI platform is designed to automate the healthcare industry’s most repetitive tasks, freeing up administrators to work on higher-level ones. The platform automates everything from eligibility checks to un-adjudicated claims and data migrations so staffers can focus on providing better patient service.

Source:



Healthcare

“ML for a better patient journey.”

Description: CloudMedX uses ML to generate insights for improving patient journeys throughout the system. Its technology helps hospitals manage patient data, clinical history and payment information by using predictive analytics to intervene at critical junctures in the patient care experience. Healthcare providers can use these insights to efficiently move patients through the system without any of the traditional confusion.

Source:



Healthcare

“Personalized healthcare plans with AI.”

Description: The Cleveland Clinic teamed up with IBM to infuse its IT capabilities with artificial intelligence. The hospital is using AI to gather information on trillions of administrative and health record data points to streamline the patient experience. This way they can personalize healthcare plans on an individual basis.

Source:



Business

“SEO friendly, high-quality article generating software.”

Description: Use this cutting-edge AI text generator to write stories, poems, news articles, and more. It is flexible as it has a large range of results depending to the request, and therefore explores creativity beyond human capabilities with certain speed.

Source: <https://www.theverge.com/11dtr/2019/5/13/18617449/ai-text-gen>



Safety & Security

“Algorithm to detect benefit fraud.”

Description: Computersystem SyRI was stopped after verdict. It used an algorithm to detect benefit fraud. According to the judge, the digital system interfered too deeply with the private lives of Dutch people and therefore violated human rights. Especially the mystery surrounding the system killed the State, according to a judgment of the court.

Source: De Trouw, 5 feb 2020

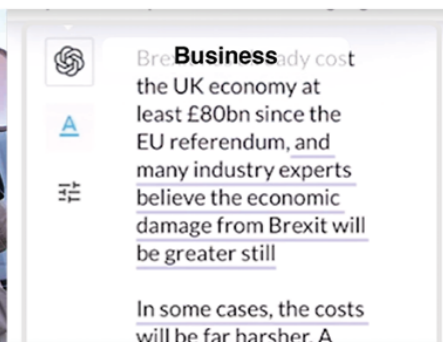


Safety & Security

“Cambridge Analytica scandal highlights need AI regulation.”

Description: A wide concern is of the creation of “data monopolies”, large multinational companies – generally American or Chinese, with Facebook, Google and Tencent all named as examples – with such a grip on the collection of data that they can build better AI than anyone else, enhancing their grip on the data sources and creating a virtuous cycle that renders smaller companies and nations unable to compete.

Source: The Guardian, 16 apr 2018



Business
Brexit could cost the UK economy at least £80bn since the EU referendum, and many industry experts believe the economic damage from Brexit will be greater still.

In some cases, the costs will be far harsher. A



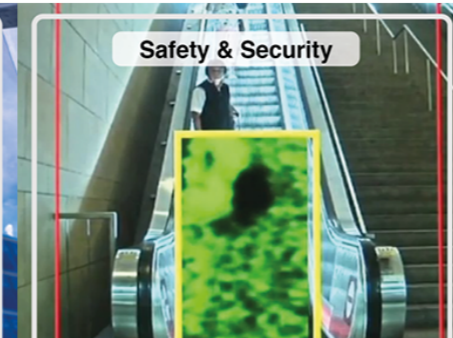



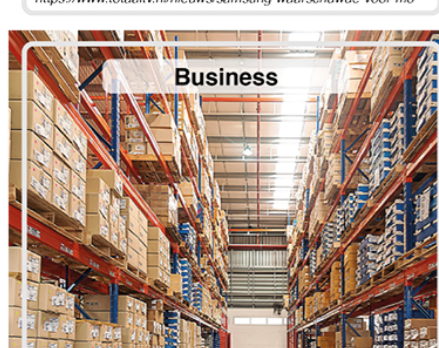

“New AI fake text generator may be too dangerous to release, say creators.”

Description: OpenAI. The creators of a revolutionary AI system that can write news stories and works of fiction – dubbed “deepfakes for text” – have taken the unusual step of not releasing their research publicly, for fear of potential misuse.

Source: The Guardian, 14 feb 2019

Ons Manifesto

Wij geloven in een **CERTIFIED** aanpak om de mensgerichte AI adoptie te kunnen bereiken. We hebben te maken met complexe problemen met veel onzekerheden, waar geen pad is dat ons de weg wijst. Vanwege het adaptieve karakter van AI en het feit dat de behoeften van de samenleving naar loop van tijd veranderen, moeten we al doende leren begrijpen welke diensten we moeten aanbieden op het gebied van veiligheid en recht. Daarom moeten we onze eigen weg uitstippelen, geïnspireerd door de agile-methodiek. Innovatie is onze plicht, de mens centraal stellen in het gebruik van technologie is ons doel.

 <p>Safety & Security</p> <p>“Identifying crime suspects on street cameras.”</p> <p>Description: Alibaba-backed SenseTime software matches crime scene offenders to criminal databases in seconds to remove threats off streets. They parse data from thousands of live camera feeds and can be used by police to “track everything from vice and accidents to suspects on blacklists.”</p> <p>Source:</p>	 <p>Safety & Security</p> <p>“Identify criminals’ faces via CCTV.”</p> <p>Description: One Smart Labs, is integrating AI into existing CCTV systems to identify subjects by matching faces with a database of millions of photos (including mug shots and driver’s licenses). As this tool could raise fears among ordinary citizens of privacy invasion, they continually test their tools’ algorithms to identify false positives and matches and minimize bias via deep learning.</p> <p>Source:</p>	 <p>Safety & Security</p> <p>“Detecting threats in facilities via object recognition.”</p> <p>Description: Deep Science AI is working on object recognition within its analytics-powered platform. It supports remote monitoring of facilities such as offices and retail stores to detect threats including weapons, masks, intruders, and fire. A human analyst monitors feeds from the facilities’ security cameras and immediately contacts the business and authorities when a threat emerges.</p> <p>Source:</p>
 <p>Business</p> <p>“TV voice recognition to target advertisements.”</p> <p>Description: Samsung smart tv uses AI to create detailed userprofiles. Active Voice Amplifier recognizes background noise in the space automatically and adjusts voice clearance. Their objecting Tracking Sound+ (OTS+) creates a 3D audio effect geluidseffect. Agreeing to the terms, they are able to overhear conversations. Their ‘Voice Recognition’ can capture sensitive info which lead to fuss about lack of privacy.</p> <p>Source: https://www.totaaltv.nl/nieuws/samsung-waarschuwde-voor-mo</p>	 <p>Business</p> <p>“AI assistant streamlines recruitment process.”</p> <p>Description: Mya Systems has developed a digital assistant that can streamline the recruitment process in various ways, view resumes, gather candidate information, and schedule job interviews. Previously, recruiters turned out to be overwhelmed by “mindless” work. They believe that these chatbots help with the bureaucratic and repetitive aspects of the recruiting process.</p> <p>Source: https://techcrunch.com/2019/03/08/polyai/</p>	 <p>Business</p> <p>“Conversational AI agents for personal customer service.”</p> <p>Description: A Startup Poly created AI agents that can independently take over and handle customer conversations, so that their human colleagues have more time for complex problems that require empathy and emotional intelligence. Their agents are built around a framework for modelling context, which means they can hold long conversations and remember all pieces of information that users had previously shared.</p> <p>Source: https://techcrunch.com/2019/03/08/polyai/</p>
 <p>Business</p> <p>“Predicting demand for inventory management.”</p> <p>Description: Lucky Brand uses advanced analysis to optimize inventory management. This lifestyle clothing brand uses the Cellect AI platform to analyze inventory in distribution centers and allocate items to stores in real time. The platform allows employees to work smarter and faster, saving them a lot of time</p> <p>Source: https://techcrunch.com/2019/03/08/polyai/</p>	 <p>Healthcare</p> <p>“More accurate cancer diagnosis with AI.”</p> <p>Description: PathAI is developing ML technology to assist pathologists in making more accurate diagnoses. The company’s current goals include reducing error in cancer diagnosis and developing methods for individualized medical treatment.</p> <p>Source:</p>	



We genereren innovatieve oplossingen, gesterkt door **Creativiteit**. We nodigen uit om te spelen, met een open en positieve mindset, om in co-creatie tot nieuwe ideeën te komen.



We vermijden reactief beleid en laten ons niet leiden door de waan van de dag. Wij werken visie gedreven en besteden structureel tijd aan de **Focus op Toekomst**. We streven ernaar om te inspireren en gebruiken onze visie als basis voor beslissingen op korte en lange termijn.



We genereren draagvlak, te beginnen met **Empathie** met de betrokken partijen. We onderzoeken latente behoeften door verschillende standpunten over de context aan te duiden, voordat we besluiten onze ideeën in praktijk te brengen.



We bieden een veilige omgeving voor mogelijke mislukkingen en nemen **Iteratie** mee in het beleidsontwikkelingsproces. We leveren tussentijdse resultaten in korte cycli door ‘Minimum Viable Policies’ in te voeren en contextuele microregels parallel te testen.



We genereren leervermogen, behouden oog op proces en resultaat, door te **Reflecteren** op het systeem als geheel. Wij omarmen kritiek, aangezien een klacht een geschenk is, met rijke inzichten van ervaringen.



Om de knowhow te verkrijgen, zijn onze keuzes **Ervaring-gedreven**, met structurele expertpanels, die beleid definiëren met het ‘hele systeem in de kamer’. We valideren onze aannames en nodigen de quadruple helix (overheid, bedrijf, academia en maatschappij) uit aan te sluiten.



Begrip wordt bereikt door onze **Transparantie**. Onze handelingen moeten voor iedereen begrijpelijk zijn, dus mensen meenemen in onze gedachten betekent open zijn over onze experimenteerprocessen en een lage drempel bieden om samen betere resultaten te ontwikkelen.



We geven betekenis niet alleen met woorden, maar verwijzen ook naar experimenten, waardoor rijke discussies mogelijk zijn die tastbaar zijn. We overbruggen talen door middel van **Demonstraties**, waarbij we experimenten omkaderen als een middel om de maatschappelijke dialoog over de toekomst die we wensen te faciliteren, door transparantie voor het publiek te brengen.



We waarderen **Inclusiviteit**, mogelijk gemaakt door open dialogen bij de start van onze experimenten. We maken face-to-face afstemming mogelijk in verticale, horizontale, cross-silo en interdepartementale teams, waardoor de barrière tussen beleid en uitvoering wordt verlaagd.

BELANGHEBBENDE

ROL

- Bedrijf opschalen
- Partner worden
- Intellectueel leiderschap

ACTIES

- waarden communiceren
- klant uitdagen
- voorlopen op competitie

WAARDEN

- Ambitie
- Macht

POWER

- rol overheid overnemen
- sneller ontwikkelen dan regels zijn opgesteld

VRAGEN

- Hoe kunnen we een partner worden?
- Wat is onze rol in het experiment?
- Welke data is al klaar voor gebruik?

BIG TECH

ROL

- Bedrijf opschalen
- Partner worden
- Intellectueel leiderschap

ACTIES

- waarden communiceren
- klant uitdagen
- voorlopen op competitie

WAARDEN

- Ambitie
- Macht

POWER

- rol overheid overnemen
- sneller ontwikkelen dan regels zijn opgesteld

VRAGEN

- Hoe kunnen we een partner worden?
- Wat is onze rol in het experiment?
- Welke data is al klaar voor gebruik?

ONDERNEMER

ROL

- Bedrijf opschalen
- Leveren van werkende technologie

ACTIES

- opbouwen netwerk
- leveren meerwaarde
- kansen identificeren

WAARDEN

- Ambitie
- (Economische) vrijheid

POWER

- implementatie afbreken

VRAGEN

- Hoe houden we ons aan wat wettelijk mag?
- Hoe kunnen we van dienst zijn?
- Hoe kunnen we duurzaam / winstgevend zijn?

FILOSOOF

ROL

- Reflecteren op geheel
- Zoeken naar bron van klachten

ACTIES

- samenhang beschrijven
- wetenschappelijk onderzoeken
- standpunten uiteenzetten

WAARDEN

- Nieuwsgierigheid
- Relevantie

POWER

- opinie delen
- debatteren

VRAGEN

- Wat is de samenhang tussen de elementen?
- Wat is het effect hiervan op de toekomst?
- Wat is de bron van het fenomeen?

JOURNALIST

ROL

- Ombudsman
- Zoeken naar waarheid en sensatie

ACTIES

- bias herleiden
- kritische vragen stellen
- rapporteren aan media

WAARDEN

- Eerlijkheid
- Kritisch

POWER

- selectief vergrootglas
- ophef door de toon

VRAGEN

- Wat wil en moet de samenleving weten?
- Hoe de boodschap communiceren?
- Wat kan er fout gaan?

BURGER

ROL

- Perspectief bieden
- Reflecteren op dienstverlening

ACTIES

- verantwoordelijkheid peilen
- bewustzijn creëren voor effect op leven van het individu

WAARDEN

- Veiligheid
- Privacy
- Gelijkheid

POWER

- in opstand komen

VRAGEN

- Wat gebeurt er en hoe leer ik dit kennen?
- Hoe wordt mijn mening hierin gewaardeerd?
- Hoe beïnvloedt dit mijn persoonlijke leven?

PRIVACY ACTIVIST

ROL

- Debat mogelijk maken
- Stress urgentie
- Vertegenwoordig onbevoegden

ACTIES

- media aandacht creëren
- machtigen onbevoegden draagvlak creëren

WAARDEN

- Rechtvaardigheid
- Betrouwbaarheid

POWER

- actie negatief belichten
- stekker uit proces trekken

VRAGEN

- Hoe wordt onze mening overwogen?
- Hoe kan het onze mensenrechten schenden?
- Welke tussenkomst is nodig om ons te horen?

RAADSADVISEUR

ROL

- Conglomerat
- Beslisbevoegdheid
- Mandaat bieden

ACTIES

- minister informeren
- bestuur organiseren
- besluiten prioriteiten

WAARDEN

- Consistentie
- Controle

POWER

- geen akkoord geven
- bezuinigen

VRAGEN

- Hoe maak je er een consistent verhaal van?
- Hoe kunnen we dit snel organiseren?
- Hoe houdt je het onder controle?

POLITIEK LEIDER

ROL

- eind-verantwoordelijk woordvoerder
- verantwoordelijkheid afleggen

ACTIES

- verhaal samenstellen
- kritiek omarmen
- urgentie benadrukken

WAARDEN

- Verantwoordelijkheid
- Ambitie

POWER

- prioriteiten verleggen
- besluiten nemen over anderen

VRAGEN

- Wat wil de samenleving (weten)?
- Hoe breng je de boodschap over?
- Hoe moet ik inspringen bij een incident?

RECHTER

ROL

- Grenzen definiëren
- Bepalen van strafbare feiten

ACTIES

- op feiten baseren
- interpreteren wetten
- concrete voorbeelden geven

WAARDEN

- Gelijkheid
- Rechtvaardigheid

POWER

- schuldig verklaren
- projecten stopzetten

VRAGEN

- Wat is wenselijk en wat is stafbaar?
- Welke regels komen hiervoor in aanmerking?
- Wat zijn concrete voorbeelden ter interpretatie?

HANDHAVER

ROL

- Misdaad bestrijden
- Behoeften van uitvoering benadrukken

ACTIES

- snel handelen bij incidenten
- interactie behoeften definiëren

WAARDEN

- Dienstbaar
- Veiligheid

POWER

- vooroordelend handelen
- keuze wel/niet ingrijpen

VRAGEN

- Wat werkt het beste voor mij?
- Hoe beïnvloedt het mijn dagelijkse werk?
- Welke waarde biedt het in mijn dienstbaarheid?

ONTWERPER

ROL

- Talen overbruggen
- inspireren
- creatief oplossingen bieden

ACTIES

- visuele referenties creëren
- proces expliciet maken
- speelsheid oproepen

WAARDEN

- Creativiteit
- Empathie

POWER

- nudgen op basis van ontwerpkeuzes

VRAGEN

- Wat zijn ieders onderliggende behoeften?
- Hoe zou superman het probleem oplossen?
- Hoe kan een oplossing concreet vorm krijgen?

(DATA) SCIENTIST

ROL

- Haalbaarheid testen
- Technische oplossing bouwen

ACTIES

- prototype creëren
- haalbaarheid toetsen
- mogelijkheden blootleggen

WAARDEN

- Exploratie
- Creativiteit

POWER

- als enige begrijpen hoe het systeem werkt
- ontwerpkeuzes maken

VRAGEN

- Wat is technisch haalbaar?
- Welke data is beschikbaar?
- Hoe het probleem technisch op te lossen?

BELEIDSMAKER

ROL

- Reflecteren en rechtvaardigen
- Draagvlak verkrijgen
- Beleid opstellen

ACTIES

- minister informeren
- informatiestroom organiseren
- vertalen wetten

WAARDEN

- Deskundigheid

POWER

- selectief betrekken

VRAGEN

- Wat moet er worden gecommuniceerd en hoe?
- Welk beleid voldoet (niet)?
- Hoe het proces organiseren en bewaken?

AFDELINGSHOOFD

ROL

- Facilitatie rol
- Beslisbevoegdheid
- Mandaat bieden

ACTIES

- akkoord geven op uitvoering projecten
- besluiten prioriteiten

WAARDEN

- Efficiëntie
- Gemak

POWER

- geen akkoord geven
- bezuinigen

VRAGEN

- Wat te regelen om het in gang te zetten?
- Hoe krijg je iedereen aan boord?
- Hoe weegt het voordeel op tegen de kosten?

CRIMINEEL

ROL

- Zwakte plek zoeken
- Tegendraads werken
- Stelen/Geweld plegen

ACTIES

- de wetten overtreden
- systeem hacken

WAARDEN

- Vrijheid
- Macht

POWER

- angst zaaien
- misbruik maken van systeem

VRAGEN

- Hoe kan ik het systeem voor mijn eigen voordeel misbruiken?
- Welk effect zouden mijn acties hebben?

UITLEG

WAT: Dit kaartdeck bestaat uit 15 belanghebbenden, kaarten en 3 blanco kaarten. Elke kaart vertegenwoordigt een belanghebbende, wiens rol, acties, waarden, macht en vragen worden beschreven. Het doel van dit kaartdeck is om de verschillende standpunten van de belanghebbenden aan het begin van een AI-experiment te identificeren en te leren begrijpen.

WANNEER: Dit kaartdeck is bedoeld om te worden gebruikt in een vergadering met een ervaringspanel, gepland aan het begin van een AI experiment.

HOE: Leg de kaarten in het midden van de tafel. Identificeer wie aanwezig is bij de vergadering en welke partij ontbreekt. In het geval dat er iemand is die niet op het kaartspel vertegenwoordigd is, maar erbij betrokken zou moeten zijn, neem dan een blanco kaart en laat de persoon deze invullen.

WAAROM: Inclusiviteit is de sleutel om een duurzame innovatie te creëren. Het combineren van expertise op de verschillende gebieden die de impact van een dergelijke applicatie zullen ervaren, zal een rijke dialoog opleveren en helpt om samen oplossingen te creëren voor slechte problemen.

WIE: Het ervaringspanel moet bestaan uit minimaal 8 van de vertegenwoordigers van de Quadruple Helix, waaronder: bedrijfsleven, overheid, academische wereld en burgers. Wanneer de vertegenwoordigende partijen niet allemaal de vergadering kunnen bijwonen, kan de niet-vertegenwoordigde kaart alsnog worden besproken.

WAAR: Deze bijeenkomst dient plaats te vinden in een creatieve en veilige omgeving, bij voorkeur buiten het reguliere kantoor.

MINDSET

Een staakt is een geschenk (omarm de waarde van feedback)

Werk naar consensus (zoek naar gedeelde waarden)

Voorbeelden overbruggen onbegrip (beschrijf a.d.h.v. referentiepunten)

Openheid biedt openbaring (durf te 'falen')

DEFINIËER CANVAS

45 min 

In dit canvas gaan we op zoek naar een consensus over een wenselijke AI toepassing. We gaan hier dan ook proberen te ontdekken wat ieder wenselijk acht in de huidige casus. Aangezien we niet een oneindige lijst kunnen maken van het spectrum wat we wenselijk vinden, starten we met het duiden in concrete voorbeelden wat we in ieder geval NIET wenselijk vinden. Bekijk de richtlijnen voor verantwoord AI: Wat zou in deze casus niet wenselijk of toegestaan mogen worden? Plak post-its van de gedachtenspinsels in het betreffende gebied. Start individueel 5min, bespreek vervolgens de inzichten tezamen en cluster vergelijkbare gedachten. Geen overeenstemming? Plaats het in het grijze gebied.

onwenselijk ← **grijs gebied** (geen consensus over) → wenselijk

Case toelichting (kopie):

.....

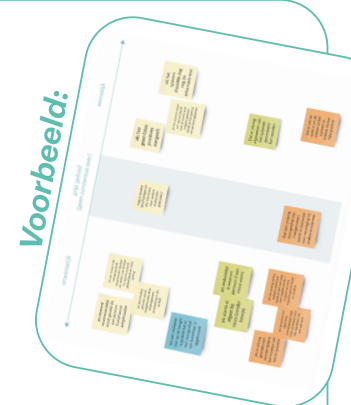
.....

.....

.....

.....

Voorbeeld:



Richtlijnen voor verantwoord AI:



Reflectie:

Welke vereisten kunnen we afleiden van de voorbeelden?

IDEE CANVAS

45 min 

Hoe kan je die zorgen oplossen met beleid? In dit canvas gaan we brainstormen over oplossingsrichtingen. De eerste verkenning van ideeën leiden mogelijk tot nieuwe inzichten. Kopieer de probleemstelling hier en formuleer het in de vorm: "Hoe kunnen we..."

Probleemstelling:

Hoe kunnen we

.....

.....

.....

1e gedachten:

Plaats hier je eerste gedachten (purge)

Brainstorm Ronde 1:

Sta open voor "radicale" ideeën geïnspireerd op de triggerkaartjes. Plak elk idee op een aparte post-it in dit vlak:

Brainstorm Ronde 2:

Haal inspiratie uit de "radicale" ideeën door ze te vertalen naar een toepasbaar idee. Plak elk idee op een aparte post-it in dit vlak:

Wat weten we nog niet?

Dit is de parkeerplaats voor hypothesen/vragen/ideeën die nog verder onderzocht moeten worden:

Reflectie:

Wat haal je uit deze ideeën? Wat zijn de top 3 meest kansrijke ideeën die je wilt testen? Plak ze hieronder op een post-it.

MINIMUM VIABLE POLICY

30 min 4

Dit canvas kan worden gebruikt om de inspanningen en middelen te besteden waar ze het meest tellen in een onzekere context. Het biedt structuur om je te concentreren op het probleem dat je probeert aan te pakken en identificeert en valideert basisaannames voordat het beleid definitief wordt gemaakt.

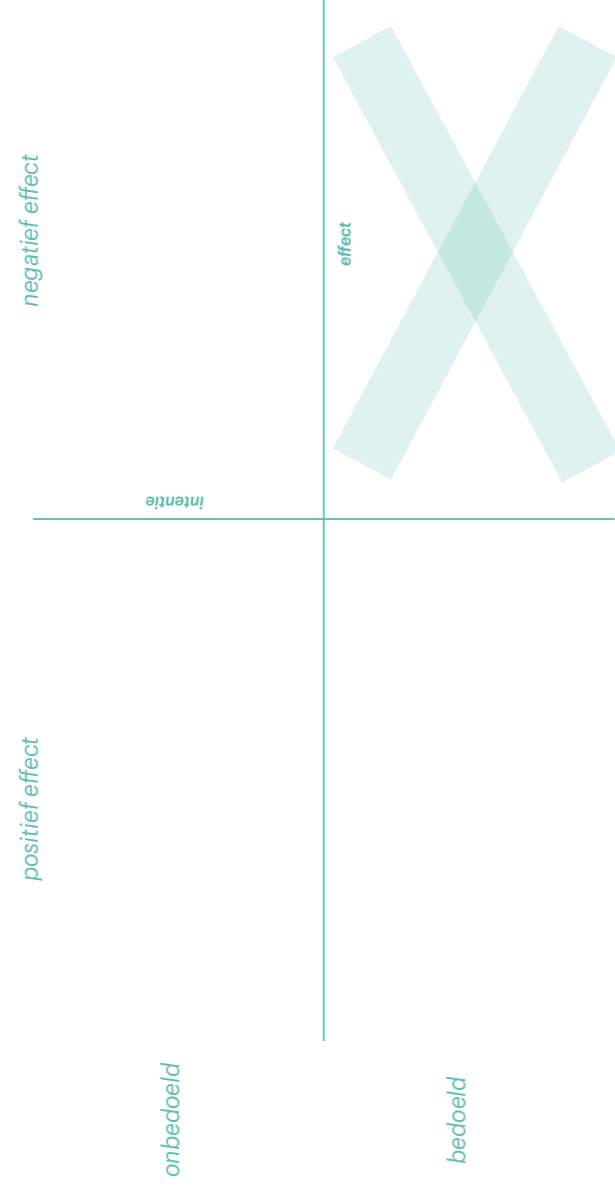
Fact-check:

Beschrijf hier welke beleidsopties zijn gekozen en beantwoord de vraag: Waarom deze t.o.v. de alternatieve wegen?

Uit de eerdere canvassen zijn vereisten ontstaan die opgevolgd dienen te worden om het AI-systeem binnen de wenselijke kaders te houden. Beschrijf de eerste aanzet tot het nieuwe beleid hieronder. (kanttekening: het gaat om een Minimum Viable Policy, dus beschrijf de minimale randvoorwaarden):

Effecten van de Minimum Viable Policy:

Wat voor effect verwacht je dat het beleid zal hebben? Naarmate de tijd kunnen ook onbedoelde negatieve effecten ontstaan. Verken de gebieden door post-its met voorbeelden in onderstaande grafiek te plaatsen:



Formuleren van Minimum Viable Policy:

Doel:

Gekozen middel:

Redenatie achter gekozen middel:

Randvoorwaarden voor het middel:

TEST CANVAS

30 min 5

Dit canvas beschrijft de testopzet van de Minimum Viable Policy. Om het beleid te toetsen dienen als eerst de gemaakte aannames gevalideerd te worden. Onderstaande vlakken helpen bij het bepalen van de testopzet. Om te zorgen dat er verantwoordelijkheidsgevoel ontstaat dient dit canvas getekend te worden door de aanwezigen bij het opstellen van de test. Na 2 weken dient een reflectie bijeenkomst gehouden te worden om de leerervaringen te delen.

Opzet test:

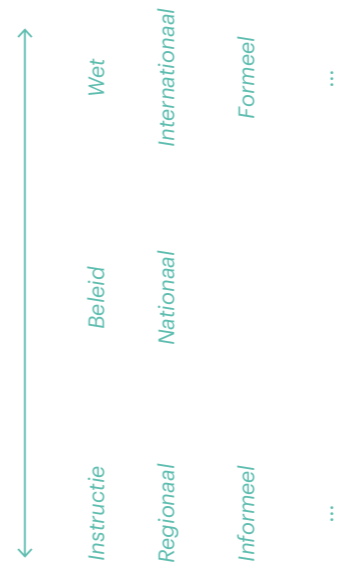
Beschrijf hier drie beleidsopties die je in een test tegen elkaar wilt afwegen:

Wanneer is de test een succes?

Hoe ga je dit meetbaar maken?

Selecteer toepassingsgebied:

Op welk niveau dienen de beleidsopties getest te worden? Omcirkel wat van toepassing is en vul lege gebieden eventueel aan waar nodig:



Testomgeving:

Beschrijf hier de testomgeving in detail: Denk hierbij aan met welke partij wordt samengewerkt, welke middelen nodig zijn, over welk tijdsplan de test wordt uitgevoerd en hoe er gemeten wordt

Wie:

Wat:

Waar:

Waarom:

Wanneer:

Hoe:

Acties Tijdlijn:

Geef hier op de lijn de stappen aan hoe de test uitgevoerd gaat worden in komende 2 weken:

Verantwoordelijkheid

Wie wordt verantwoordelijk voor de uitvoering van de test?
Naam:

Wie wordt de beleidsverantwoordelijke?
Naam:

Voorbeeld:



APPENDIX G



Wat zou een toekomstige krantenkop zeggen?



Hoe zou je proberen het systeem te misbruiken?



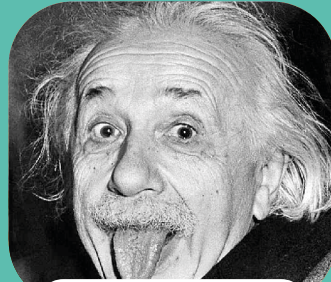
Hoe zou je het in 2050 aanpakken?



Wat als je toegang had tot alle informatie?



Wat als je het tegenovergestelde zou doen?



Hoe zou je het in 1950 hebben aangepakt?



Wat als je een oneindig budget zou hebben?



Wat als je een superkracht zou hebben?



Oma Wijsheden: Wat zou oma adviseren?



Wat zou je doen als je maar 24 uur zou hebben?



Hoe zou jouw superheld het oplossen?



Hoe zou Trump het oplossen?



Hoe wordt het opgelost in een andere industrie?



Wat zou je doen als je je acties kan terugdraaien?