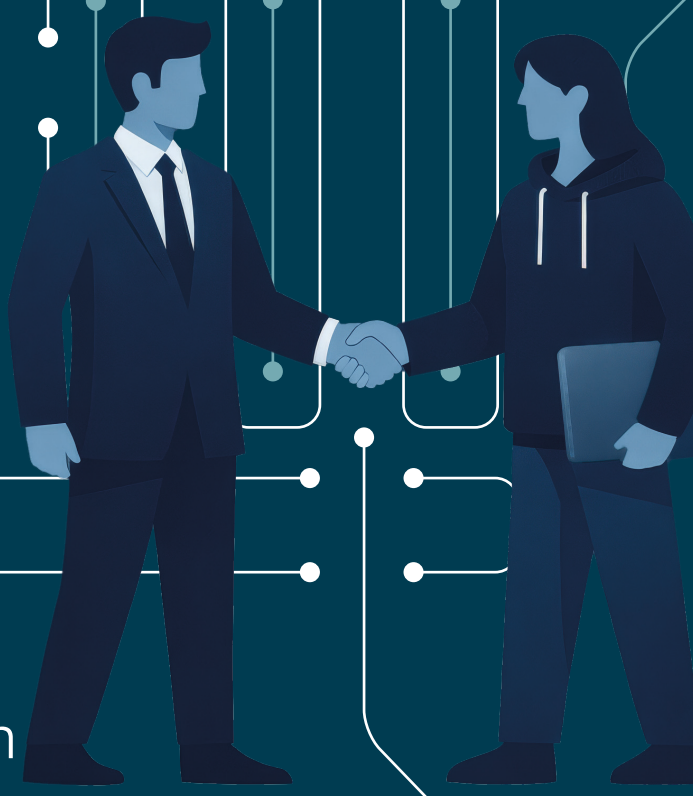
PUBLIC ORGANIZATIONS IN TRANSITION

AI, Professionals, and Organizational Change



Floris van Krimpen

PUBLIC ORGANIZATIONS IN TRANSITION

AI, Professionals, and Organizational Change

Floris van Krimpen

Public Organizations in Transition

AI, Professionals, and Organizational Change

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof.dr.ir. H. Bijl
chair of the Board for Doctorates
to be defended publicly on
Friday 6 March 2026 at 12:30

by

Floris Justus VAN KRIMPEN

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof.mr.dr. J.A. de Bruijn	Delft University of Technology, promotor
Prof.dr. M. Arnaboldi	Politecnico di Milano, promotor
Dr. H.G. van der Voort	Delft University of Technology, copromotor

Independent members:

Prof.dr.ir. N. Bharosa	Delft University of Technology
Prof.dr. W.W. Veeneman	Delft University of Technology
Prof.dr. G. Azzone	Politecnico di Milano, Italy
Prof.dr.ing. A.J. Klievink	Leiden University
Prof.dr. M.A. van der Steen	Erasmus University Rotterdam

The doctoral research has been carried out in the context of an agreement on joint doctoral supervision between Politecnico di Milano, Italy and Delft University of Technology, the Netherlands.

Key words: artificial Intelligence, professionalism, organizational change, digital transformation, co-creation, generative AI, machine learning

Printed by Proefschriftspecialist | proefschriftspecialist.nl
Layout and design: Timo Wolf Kamp, persoonlijkproefschrift.nl

ISBN: 978-94-93483-76-7

An electronic version of this dissertation is available at: http://repository.tudelft.nl

"...what matters in life is not whether we receive a round of applause; what matters is whether we have the courage to venture forth despite the uncertainty of acclaim."

– Amor Towles, A Gentleman in Moscow

Table of Contents

Preface

While I started third grade of primary school unable to read, at the end of the year I could almost read flawlessly at the highest AVI level. Apparently, I had a knack for reading. Throughout my life this ability has stayed with me. It supported me in developing a strong love for reading. It started with all kinds of children's books and the wish to read all the books of the famous Dutch 'De Kameleon' series. Who doesn't want to read about crazy adventures by two mischievous and adventurous young brothers with a very fast boat? Today I not only read about adventures anymore, as my interests have broadened.

At some point during my reading journey I realized I wanted to write a book myself. If they can do it, why couldn't I? When the opportunity to pursue a PhD presented itself I had doubts. Starting such a project, with such a big timeline, is that really a good idea? Ultimately, I decided to go for it. I believe that one of the key reasons behind that decision is my love for reading and the wish to write a book myself.

Then came the next question: What should I write about? For me, like for many others colleagues who started during the same time, the *Toeslagenaffaire (*Dutch childcare benefits scandal) was a trigger. This world of algorithms, risk-profiling, and societal impact seems interesting. The affair impacted many people in a very negative way. Could I contribute to this topic in a way that positively effects society?

For various reasons my research and I took a slightly different direction. The focus shifted from broader society and citizens towards organizations and the professionals within them. I also avoided an overly strong focus on the values at stake. I observed that organizations were starting to embrace AI. But what happens within those organizations? And how are the professionals within dealing with this change? In essence, the bottom line has not changed much.

I hope my research brought together in this book positively affects people working with AI within public organizations. Hopefully my research helps them understand how their work will change and how they can deal with these changes. This, in turn, should contribute to responsible AI-adoption with positive effects on society and its citizens.

Acknowledgements

In the first year of my PhD, I doubted many times whether I would even write these acknowledgements. Many times I asked myself the questions: Should I even be doing a PhD? Is this the right path for me? Being in the midst of the Covid19 pandemic probably played a role. I had barely started my PhD when we went into the first lockdown. It felt lonely. Over the years doing a PhD, I have learned that it can stay lonely. In the end, doing a PhD is an undertaking you do by yourself. Only you have the capacity to bring the project forward. But, luckily, simultaneously I have also experienced the other side. While doing a PhD is something you do by yourself, it is also impossible to finish without the support of many – and when I say many I mean maaaany - others. This is the place where I express my gratitude to all of you. Thank you for making the PhD journey so much more fun.

First, I think of *Hans*. One of my promotors. Thank you for being the promotor and person that you are. You are one of the reasons that led me to do a PhD. After my Master you asked me to work for you as a research assistant. During that time you asked me whether I would be interested in doing a PhD, because you thought it could be something for me, probably based on the excellent literature reviews I handed in every Friday :). While I initially decided to go in a different direction, we always stayed in touch. Eventually, we decided to go for it. And now here we are. Over the years I think I have gotten to know you a bit better. I believe this is also because you have seen that I like to get to know the person behind the role. I see you as a promotor who gives a lot of freedom, but is there when you need him. You never tell what to do, but always provide options regarding what I could do. I am really thankful for this approach, because it gave me the feeling I had autonomy. It gave me the feeling that I had the steering wheel in my hands. Your style of guidance is something that I will take with me forever.

Then, *Haiko*. Copromotor and daily supervisor. Where do I even start? Also a big thanks to you. Thank you for listening to me on the countless occasions I walked past your door and finally into your office for a question about this or that. Thank you for dropping by my office for a random chat about research, life, or any other topic, while you probably had many other far more important things to do. From the moment you entered the PhD journey things became a lot easier. Because now there really was someone I felt I could speak to on a more or less daily basis. I did not know you before the PhD, but now a bit more. I have gotten to know you as someone who can be both serious and completely not serious in the span of a short time. Someone who can be really cynical. But also someone with a lot of humour. Someone who always has an additional angle to the research we were working on. And someone who is able to deal with stubborn people like me. We had many discussions about our shared paper on a really detailed level. You took the time to really go deep. You challenged me on my assumptions. You gave me space to do it my way. You taught me to trust my instincts. I thank you for that. Through our interactions I could see more and more that I was making progress.

Michela, anche a te voglio dire un enorme grazie. As my Italian promotor you were less physically present during the PhD journey than many others. But, you have played a major role in me finishing this PhD. You were there at the start when you and *Hans* had the idea to supervise a PhD student between Delft and Milano. I want to thank you tremendously for how you have dealt with the entire process. You made me feel welcome in Milano, you were compassionate, supportive, and accepting when it turned out I would be less present in Milano than initially thought. I also want to thank you for your flexibility during the end of the PhD and for doing your best to understand the (cultural) differences that sometimes hindered the process. I have said it before, but I can be somewhat stubborn and I often want to know the 'why' behind the 'what'. This, together with what I believe are cultural and personal differences, led to us having our 'moments'. However, you never held them against me. Thank you for that.

Also, many thanks to the secretaries. Without them, getting towards this point was impossible. Especially, *Jolanda*, thank you for dealing with all my requests. Thank you for making sure that I could meet Hans when needed. Thank you for taking care of all the forms and other necessary stuff behind the scenes. Also, *Joyce* deserves a mention. Your office was across ours. Thanks for the chats, answering my questions, and everything else.

Then, my PhD colleagues in Delft. Thank you all for making the PhD journey so much more fun. Like said, doing a PhD is a major undertaking. Only those in the same boat will truly understand what it's like. Therefore, having such colleagues, with similar challenges and issues is of utmost importance. I think of *Arnoud* and *Ana*, my office roommates. Thank you both for the laughs, the serious talks, the walks, the lunches, the coffees, the dinners, and whatnot. I am happy to have shared this path with you. But, you were not the only ones. The group of colleagues became bigger over time. J*orge*, we had our lunches, talks, walks, and coffees. I enjoyed them a lot. *Marie*, thank you. Not only for the shared lunches and the sweets that I stole from your and *Jorge's* office , but also thank you for dropping by in Milan, on your way to a conference in Florence. *Sem*, thanks for the lunches, walks and especially talks about our shared topic. We both studied AI in public organizations. It was always nice to talk with you about our ideas and perspectives. We even shared an apartment while at a conference in Leuven. *Antonia*, thank you as well. We did not speak that much. But, when we spoke it was always good. Others who deserve special mentioning are *Fabienne, Inigo, Tonny, Michael, Dewant, Janneke, Jeroen,* and *David*. You all played a role, big or small.

The same goes for my colleagues in the *AI Futures Lab* colleagues. I will not mention you all by name. However, I am thankful to have been a part. It gave me perspectives I hadn't seen before on topics I did not know much about yet. The same goes for a few other colleagues from the O&G Section. *Mark,* you are not at the TU Delft anymore, but you have been missed. Dropping by our office to have a little chat about Feyenoord or similar topics always contributed positively to my day. *Toyah,* we

have been office mates for a very short while. But you have helped me whenever I have asked. Also when we weren't office mates anymore.

From the Netherlands we move towards Italy. *Remaps* group. Thank you for being so welcoming to a random Dutch guy from Rotterdam who tries to be a bit of an Italian. I will not mention everyone by name, because I am afraid I will forget a few. Let's just say I enjoyed being a part of the group a lot. Whether it was the lunches, some drinks or just a coffee, I look back at it with much gratitude. However, there are some that need to be specially mentioned. First and foremost, *Lorenza*. You saw this one coming, didn't you? My time in Italy and working on the shared paper with *Michela* would have been tremendously different if it weren't for you. *Grazie mille*! You helped me make sense of how things were done in Italy, you were strict on me whenever I mispronounced Italian words, you acted as a liaison between me and *Michela* whenever I did not understand what was expected of me, you carried out the interviews for our shared paper. The list could go on. Again, *grazie!* It would not have been the same without you. *Alessandro* and *Romain. Alessandro,* I still remember you waiting for me the first day of my first stay in Italy. We shared some breakfast. Thank you for the talks about football and life, the drinks, the playing football, and just for being a kind guy. Hopefully we will meet again someday, preferably in Sicily. *Romain,* thank you as well. You were welcoming and together with *Alessandro* welcomed me to the football group. You both have made my stay a lot more fun.

We are nearing the end. But we are not there yet. I want to thank everyone who participated in this research or that have helped me and my co-authors find people who want to participate. Without all of you this book would not have existed. Especially *Stephanie,* thank you. We have had many chats about the research and the implications for the ILT. You were always constructive and made time to exchange perspectives. In that way, you have played a big role in forming my research and ideas.

Frans, my boss at Bisnez, my current employer. Thank you for giving me space to finish my PhD. Especially in the beginning, when I started,

some additional work was necessary. Without that flexibility and a bit of freedom, it could definitely have taken a bit longer to finish.

My family, especially mum *Riena*, my dad *Kees*, *Anne, Friso*, and also *Carel*. Dad, thanks for listening to me whenever it was needed. You taught me to be ambitious, and to try to make a positive impact. You partly inspired me to do a PhD, since you have done one yourself. Mum, you listened to me countless times whenever things got tough, and shared experiences that you heard from PhDs within the Erasmus Medical Center. You encouraged me to keep on going, but never tried to reduce the struggle doing a PhD sometimes was for me. We had our walks during lunchtime, when I still lived at the Coolhaven. I am thankful for those and everything else. *Friso,* thanks for being a supportive younger brother. I don't think you had a necessarily big role in the PhD perse, but most definitely in my life. Thanks for the jokes, the talks, and for your support. I am happy that you are proud of being my paranymph. *Carel,* let's not beat around the bush, we have a somewhat difficult relationship. Despite this, I know you are proud of me for having done and finished a PhD. I have seen you do your own sort of PhD, building a successful company. I am proud that you have succeeded in what you set out to do.

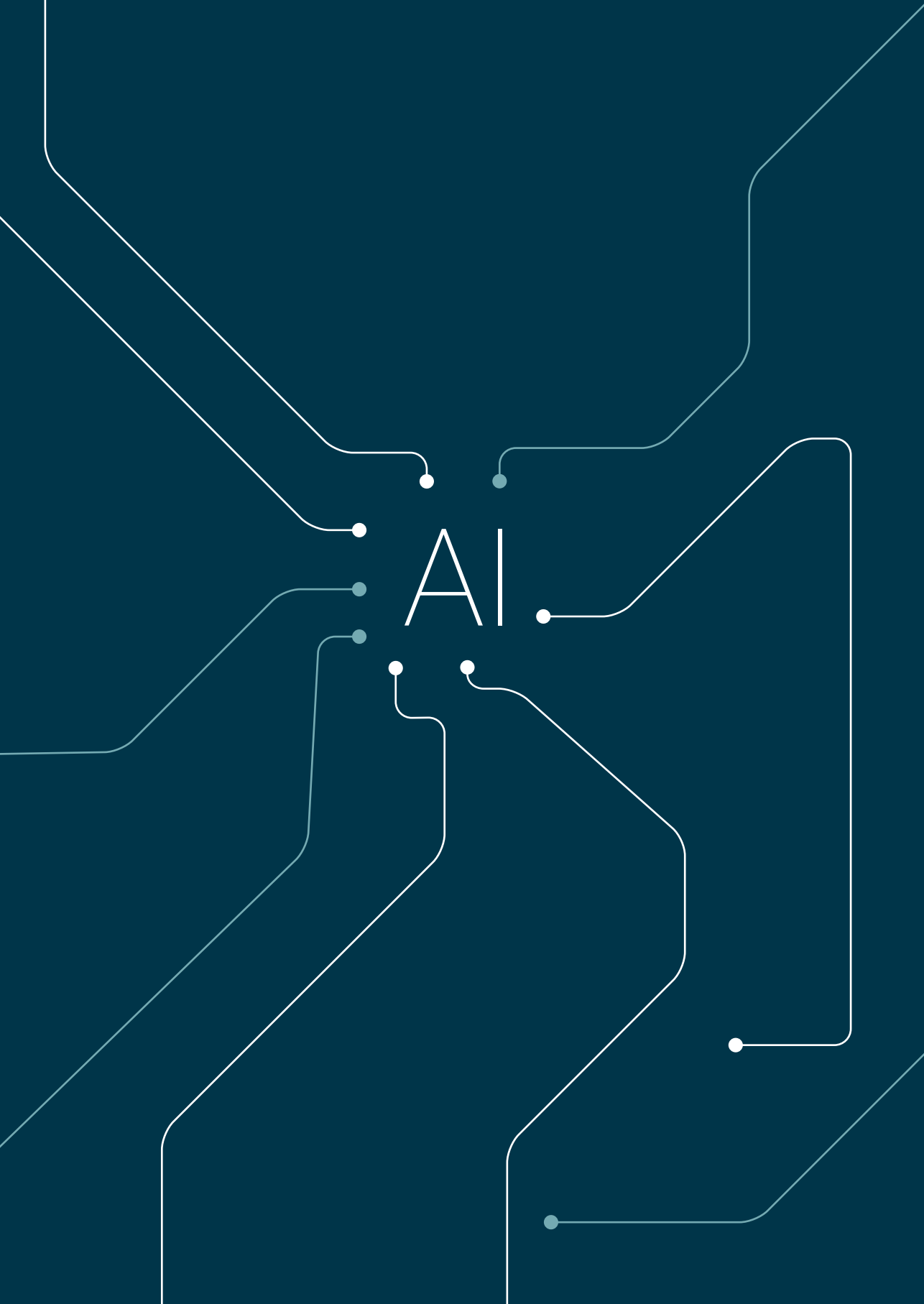
Then, my parents-in-law. *Suzanne* and *Pieter.* You were always interested in how my PhD was taking shape. You wanted to read my articles, although you said you had difficulty understanding what they were actually about. Thank you for taking an interest in what I did, for making me feel welcome, and for showing support when things became though.

Second to last, my friends. You know who you are. I will not mention everyone. However, some deserve mentioning for their special role. First *Jan*. We both started our BSc at TU Delft. We went to do different Masters. But, who would have thought that so many years later I would be defending my PhD and you would be one of the paranymphs. The circle is closed. And by the way, thank you that you kept referring to the PhD thesis as 'je werkstuk'. It made me realize to take it not too seriously at times. *Philip,* thank you for letting me work at the office. We both started

big projects more or less in the same period. You with your company, me with the PhD. I enjoyed working at the office, the drinks, and the games of ping pong. Also, thanks for dropping by in Milan. *Ralph,* the same goes for you. Thanks for coming by in Milan and just all in all being a supportive friend. You taught me that one never can have too many tiramisu's in a day. *Noah,* thanks for the friendship, and our shared lunches. I most definitely owe you a few. Especially in the early years you treated me many times, as I was a not-so-wealthy PhD student at the time. *Sebas,* we both stayed in Milan at the same time. We shared beers, lunches, and some dinners. We watched football matches and had some good nights out. Thanks for keeping me company. *Hidde,* I want to mention you as well. Especially in the last months of the PhD you were on my mind a lot. Not only because of the great example you gave defending your PhD, but mostly because of the terrible loss of *Elisabeth.* I still remember how proud she looked at you at your defense. And I am sure she would be extremely proud on how you are dealing with her loss. I am sure I am! Finally, *Alberto.* My great Italian friend. Life can be strange. We randomly met when I was on an exchange semester in Milan during the Master. I believe you asked something about the class that we would have later. From that moment onwards I gained a friend. You showed me around in Verona while I was in Milan during the PhD, and we saw the most Italian football match I will ever see, with an ecstatic ending. Thank you for your friendship!

Als laatste, *Bregje* en *Juul*. Lieve *Bregje* voor jou schakel ik speciaal over naar het Nederlands. Niet omdat ik een grapje met je uit wil halen, omdat ik weet dat je soms onzeker bent over je Engels. Wel omdat mijn Engels niet toereikend genoeg is om je te vertellen hoe dankbaar ik je ben. Dankjewel voor alles! Dankjewel dat je me aanspoort om soms net dat extra stapje te doen, als ik daar zelf geen zin in heb. Dankjewel dat je me laat zien hoeveel een mens aankan. Dankjewel dat je er voor me bent als ik het lastig heb, dankjewel dat je mijn gemopper aanhoort als ik weer eens chagrijnig ben. Dankjewel dat je ongemerkt een hele hoop regelt, en dat je niet zeurt als ik dat dan weer vergeet. En zo kan ik nog wel even

doorgaan. En, last but not least, dankjewel voor onze prachtige dochter *Juul* en de manier waarop je er voor haar bent. *Juul,* het duurt nog wel even voordat je dit kunt lezen. Maar je inspireert me. Om een boefje te blijven, om te lachen, om mijn best te doen, om nieuwsgierig te blijven. Je bent nog zo klein op het moment van schrijven – ook al was je eerst nog veel kleiner – maar je hebt al zo'n grote impact. Lieverds, ik hou van jullie!


AI

Summary

Summary

AI will significantly influence the operations of public organizations. This involves much more than just executing existing tasks of these organizations more efficiently or effectively. AI adoption[1] will lead to new possibilities for public organizations that are difficult to predict at present. Consequently, AI will also have important organizational implications. The role of professionals changes, organizational structures may require adjustment, and internal and external dependencies between various actors shift.

With the rising AI adoption by public organizations, scientific interest in the organizational implications of AI adoption has also grown. Although existing literature increasingly recognizes that algorithms and organizations are intertwined, the concept 'organization' is often minimally defined, and there is still little attention given to the components that make up an organization. Additionally, the interrelationship between organization and technology has only been empirically investigated to a limited extent.

Better understanding the interrelationship between technology and organization, and attention to the components that make up the organization, are not only relevant from an organizational perspective. Examples from the past, such as the discriminatory COMPAS algorithm used by judges in the US to assess recidivism risk, demonstrate that AI can have far-reaching consequences for citizens. Given the interconnection between technology and organization, a better understanding may potentially reduce the impact on citizens as well.

1 By AI adoption I refer not only to incorporating AI within organizational boundaries, but also to the use and development of AI within public organizations. In doing so, I employ a very broad definition of AI adoption that explicitly transcends the fields of multiple research areas.

A more detailed perspective on the concept of 'organization' is needed. An organization is never a straightforward phenomenon, but can be analysed through different lenses. This dissertation offers several of these lenses and thus a more detailed perspective on public organizations that develop and use AI. If such a detailed perspective is not applied, there is a risk that the effect AI will have on public organizations, and that organizations will have on AI, will be misunderstood. Misunderstanding these effects may lead to decisions with undesirable impact in public AI adoption.

This dissertation provides a detailed perspective on AI adoption by public organizations through examination of three levels: the individual, the organization, and the context. For each level, this dissertation provides a specific analytical lens.

- At the individual level, focus is directed toward the individual professional. This professional possesses autonomy and operates and makes decisions based on experience and tacit knowledge.
- The organizational level concerns the multitude of actors involved in AI adoption - about collaboration between new and existing actors with different backgrounds, and the obstacles this creates.
- The analysis of the organizational environment builds on the contingency perspective on organizations. This involves organizational adaptation to contextual factors and the requirements that external environments impose upon public organizations adopting AI.

Through these levels and the corresponding three lenses, this dissertation answers the question: *What are the organizational implications of the adoption of learning algorithms in public organizations?*

As noted, the relationship between AI and organizations is inherently reciprocal. This dissertation answers the main research question by departing from previous research. That research highlights that AI algorithms transform existing practices within organizations, while simultaneously that existing practices within the organization may influence the way AI algorithms are adopted. This dissertation examines the two-sided relationship between AI algorithms and organizations. Through three key lenses on organizations it analyses how AI transforms organizations while existing practices simultaneously influence AI adoption.

This dissertation is based on three separate exploratory empirical studies within public organizations, a conceptual book chapter about decision-making with AI in the public sector, and a commentary on the relationship between GenAI and professionals tacit knowledge. The empirical studies correspond to the core chapters of this dissertation, chapter 4, 5 and 6. These chapters report on the multiple research questions that together answer the main research question. Those research questions are all related to the two-way process in which AI algorithms reshape the processes in which they are embedded and that existing practices within the organization may influence the way AI algorithms are adopted and used.

Chapter 4 follows from the idea that AI comes with a variety of professionals – incumbent and novel – that need to collaborate and interact with each other in order to develop AI. More specifically, often developers and end-users together create AI systems. However, the need to interact is impeded by several challenges that can limit its effectiveness, including boundaries between developers and end-users. The chapter studies the effect that interaction has on boundaries between developers and end-users in ML development through a case study. Knowledge boundaries exist concerning language, understanding, and goals, and the question is what organized interaction does to these

boundaries. The chapter concludes that on the surface interaction contributes to blurring boundaries. However, a deeper look betrays more nuanced effects. Organization interaction does indeed blur boundaries to a certain extent, but deep-rooted boundaries around language and understanding persist. Additionally, novel boundaries emerge between different types of end-users a as consequence of the involvement of only some end-users.

Chapter 5 is related to the fact that public organizations deal with various societal demands in their adoption of AI. For example, that AI use by public organizations should be explainable. To the public, but also to the professionals within the organization. Recently, interaction between actors has been associated as a mechanism for increased explainability. Given this association, chapter 5 explores the benefits and challenges of interaction on the explainability of AI models trough action research. Furthermore, it explores how these benefits and challenges occur. Interaction is mentioned as a key condition for helpful AI models. Explainability is often mentioned as a property of such models. The chapter concludes that interaction between actors involved in AI increases explainability through organizational learning. As such, it does not contribute to explainability in the more traditional technical manner, by opening up the model. Furthermore, the chapter highlights the importance of differentiating between different types of internal stakeholders and how interactions impact developer roles, stakeholder involvement, and AI development practices within public organizations.

Chapter 6 stems from the premise that AI will impact professionals, whether incumbent or novel. Specific characteristics associated with professionals, like autonomy, discretion, and tacit knowledge are expected to be influenced by AI adoption, for example by shifting discretion from incumbent professionals to new professionals or by decreasing the autonomy of incumbent professionals on the ground, especially those making decisions. Therefore, chapter 6 studies how

professionals within inspectorates expect generative AI to affect their work through a workshop and semi-structured interviews. Generative AI is expected to contribute to many repetitive tasks, but more and more also GenAI's probable impact on more complex tasks is recognized. This chapter concludes that the emergence of GenAI might lead to an even stronger importance of tacit knowledge, and a strong need for room for experimentation to facilitate learning processes of working with GenAI, and the role of in particular GenAI as a sparring partner in the daily work of oversight professionals. GenAI as a sparring partner means that a two-way relationship develops between professionals and GenAI, where these tools can also challenge and balance professionals' perspectives. This requires well-developed professionalism.

The studies bundled in this dissertation collectively help us to answer the main question: *What are the organizational implications of the adoption of learning algorithms in public organizations?* Logically, the answer to this question is multifaceted. It is clear, however, that public AI adoption has significant implications. On the one hand, this dissertation shows that public AI adoption results in increased organizational complexity and incongruence between how organizations are structured and what is needed for successful AI adoption. Phenomena that illustrate this include new boundaries (Chapter 4), discretionary dependencies between different professionals (Chapters 4 and 5), high time investments without clear returns (Chapters 4 and 5), and the necessity to acquire new skills (Chapters 4, 5, and 6). While this may not be surprising for readers already familiar with the effects of new technologies on organizations, these findings do raise questions about whether AI is worth the investment. After all, AI is often referred to as a technology that can increase efficiency and improve service delivery, but the described phenomena seem to illustrate the opposite.

However, this conclusion shows only half of the story. The other half tells a different story. As the outside world increasingly adopts AI, public

organizations cannot afford to lag behind. While adopting AI, organizations learn what is necessary for AI adoption, new professional roles emerge, and individual professionals learn to handle the daily reality of working with AI. The current incongruence described in this dissertation therefore does not imply that public organizations should not adopt AI. On the contrary, this dissertation recommends that public organizations innovate with AI.

Public organizations, and decision-makers within, are encouraged to support professionals who are eager to work with AI, to experiment with interaction, and to understand that AI adoption comes with errors, ambiguity, and imperfection. Only through experimentation can these public organizations reap the long-term benefits that AI promises. These organizations must, however, adopt AI thoughtfully, considering the values they represent, the strategic risks of excessive dependency on American GenAI solutions, and the necessity of continuous innovation. Being thoughtful requires public organizations to actively seek and monitor AI-related developments and signals. This means exercising caution while simultaneously facilitating AI adoption, remaining vigilant rather than passive. In the process of AI adoption, these organizations and their professionals will encounter various phenomena.

Phenomena such as the redistribution of discretion from existing to new professionals, the emergence of new knowledge boundaries through the waterbed effect, or an increasing variety of professionals and variety within professions. These organizations will also face persistent boundaries that prove difficult to overcome and the emergence of hybrid roles that bridge professional groups that do not understand each other. This is inherent to the transformation process surrounding AI adoption and signals that AI is increasingly becoming part of the organization. However, organizations should not throw the baby out with the bathwater in this difficult and complex process. Because in the world of AI, the professional is still very important. Without well-developed professionalism, AI cannot

be properly valued, and no counterbalance can be offered to this new technology. Thus, the foundation of the AI-driven transformation of public organizations still depends on professionals from various backgrounds and the interactions they have.

Samenvatting

AI zal van grote invloed zijn op de werkzaamheden van publieke organisaties. Daarbij gaat het om veel meer dan het efficiënter of beter uitvoeren van de bestaande taken van deze organisaties. AI adoptie[2] zal leiden tot nieuwe mogelijkheden voor publieke organisaties, die zich nu nog moeilijk laten voorspellen. Daarmee zal AI ook belangrijke organisatorische consequenties hebben. De rol van professionals verandert, organisatiestructuren kunnen aanpassing vergen, en interne en externe afhankelijkheden tussen allerhande actoren veranderen.

Met de toenemende AI adoptie door publieke organisaties is ook de wetenschappelijke interesse in de organisatorische implicaties van AI adoptie gegroeid. Hoewel bestaande literatuur in toenemende mate erkent dat algoritmen en organisaties met elkaar verweven zijn, wordt het concept 'organisatie' vaak minimaal gedefinieerd en is er nog weinig aandacht voor de onderdelen waaruit een organisatie bestaat. Daarnaast is de verwevenheid tussen organisatie en technologie empirisch slechts beperkt onderzocht.

Het beter begrijpen van de verwevenheid tussen technologie en organisatie en aandacht voor de onderdelen waaruit de organisatie bestaat zijn niet alleen relevant vanuit organisatorisch oogpunt. Voorbeelden uit het verleden, zoals het discriminerende COMPAS-algoritme dat door rechters in de VS werd gebruikt om het recidiverisico te beoordelen, tonen aan dat AI verregaande gevolgen kan hebben voor burgers. Gegeven de verwevenheid tussen technologie en de organisatie, kan een beter begrip de impact op burgers mogelijk ook verminderen.

2 Met AI adoptie refereer ik niet alleen een het opnemen van AI binnen de organisatiegrenzen, maar tevens het gebruik en de ontwikkeling van AI binnen publieke organisaties. Hiermee hanteer ik een zeer ruime definitie van AI adoptie die nadrukkelijk de velden van meerdere onderzoeksgebieden overstijgt.

Er is een gedetailleerder perspectief nodig op het concept 'organisatie'. Een organisatie is nooit een eenduidig verschijnsel, maar laat zich door verschillende lenzen analyseren. Dit proefschrift biedt een aantal van die lenzen en daarmee een meer gedetailleerder perspectief op publiek organisaties die AI ontwikkelen en gebruiken. Als een dergelijk gedetailleerd perspectief niet wordt toegepast, bestaat het risico dat het effect dat AI zal hebben op publieke organisaties en dat de organisaties zullen hebben op AI, verkeerd wordt begrepen. Het verkeerd begrijpen van deze effecten leidt mogelijk tot beslissingen met ongewenste impact bij publieke AI adoptie.

Deze dissertatie biedt een gedetailleerd perspectief op AI adoptie door publieke organisaties door naar drie niveaus te kijken: het individu, de organisatie, en de context. Op deze niveaus biedt deze dissertatie telkens een lens waardoor naar dat niveau wordt gekeken.

- Op het individuele niveau is de aandacht gericht op de individuele professional. Die professional heeft autonomie en handelt en beslist op basis van ervaring en (stilzwijgende) kennis .
- Het organisatorische niveau gaat over de veelheid aan spelers die bij AI adoptie zijn betrokken - over samenwerken tussen verschillende nieuwe en bestaande actoren met andere achtergronden, en de hindernissen die dat oplevert.
- Bij de analyse van de omgeving van de organisatie wordt gebruik gemaakt van het contingentie-perspectief op organisaties. Het gaat over het aanpassen van de organisatie aan de context, en welke eisen de omgeving stelt aan publieke organisaties die AI adopteren.

Vanuit deze niveaus en de bijbehorende drie lenzen beantwoordt dit proefschrift de vraag: *Wat zijn de organisatorische implicaties van de adoptie van lerende algoritmen in publieke organisaties?*

Zoals benoemd is de relatie tussen AI en organisaties uiteraard wederkerig. Dit proefschrift beantwoordt de hoofdvraag door eerder onderzoek als vertrekpunt te nemen. Dat onderzoek benadrukt dat AI-algoritmen bestaande praktijken binnen organisaties vernieuwen en tegelijkertijd dat bestaande praktijken binnen organisaties van invloed zijn op de manier waarop AI-algoritmen worden geadopteerd. Dit proefschrift onderzoekt deze wederkerige relatie tussen AI en organisaties. Het analyseert middels drie belangrijke lenzen hoe AI organisaties verandert terwijl bestaande praktijken tegelijkertijd de adoptie van AI beïnvloeden.

Deze dissertatie is gebaseerd op drie afzonderlijke exploratieve, empirische studies binnen publieke organisaties, een conceptueel boekhoofdstuk over besluitvorming met AI in de publieke sector, en een kort artikel over de relatie tussen GenAI en impliciete 'tacit' kennis van professionals. De empirische studies corresponderen met de kernhoofdstukken van deze dissertatie, hoofdstuk 4, 5 en 6. Die hoofdstukken beantwoorden die verschillende sub-vragen die samen de hoofdvraag beantwoorden. Deze onderzoeksvragen zijn allemaal gerelateerd aan het tweerichtingsproces waarin AI-algoritmen de processen waarin ze zijn ingebed opnieuw vormgeven en dat bestaande praktijken binnen de organisatie de manier kunnen beïnvloeden waarop AI-algoritmen worden geadopteerd en gebruikt.

Hoofdstuk 4 volgt uit het idee dat de introductie van AI leidt tot de opkomst van nieuwe professionals en dat samenwerking tussen deze nieuwe professionals en de bestaande, 'incumbent' professionals cruciaal is voor de adoptie van AI. Specifieker gezegd creëren ontwikkelaars en eindgebruikers vaak samen AI-systemen. De noodzaak van interactie wordt echter belemmerd door verschillende uitdagingen die de effectiviteit ervan kunnen beperken, zoals grenzen tussen ontwikkelaars en eindgebruikers. Het hoofdstuk bestudeert middels een case studie het effect van interactie op grenzen tussen ontwikkelaars en eindgebruikers tijden ML ontwikkeling. Er bestaan kennisgrenzen met betrekking tot

taal, begrip en doelen, en de vraag is wat georganiseerde interactie met deze grenzen doet. Het hoofdstuk concludeert dat interactie op een oppervlakkig niveau bijdraagt aan het vervagen van grenzen. Echter, beneden de oppervlakte bleken er meer genuanceerde effecten te bestaan. Georganiseerde interactie draagt inderdaad tot een zeker niveau bij aan het vervagen van grenzen, maar diep gewortelde grenzen rondom taal en begrip blijven bestaan. Daarnaast ontstaan er nieuwe grenzen tussen verschillende soorten eindgebruikers door de betrokkenheid van slechts enkele eindgebruikers.

Hoofdstuk 5 is gebaseerd op het feit dat publieke organisaties te maken hebben met verschillende maatschappelijke eisen bij de toepassing van AI. Bijvoorbeeld dat AI-gebruik door publieke organisaties uitlegbaar moet zijn. Aan de maatschappij, maar ook aan de professionals binnen de organisatie. Recentelijk is interactie tussen actoren genoemd als een mechanisme voor betere uitlegbaarheid. Gezien deze associatie onderzoekt hoofdstuk 5 de voordelen en uitdagingen van interactie op de verklaarbaarheid van AI-modellen door middel van action research. Daarnaast exploreert het hoe deze voordelen en uitdagingen tot stand komen. In de literatuur wordt interactie genoemd als een belangrijke voorwaarde voor behulpzame AI modellen. Uitlegbaarheid wordt vaak genoemd als een eigenschap van zulke modellen. Het hoofdstuk concludeert dat interactie tussen actoren betrokken bij AI de uitlegbaarheid verhoogt door organisatorisch leren. Het draagt dus niet bij aan uitlegbaarheid in de meer traditionele technische manier, door het transparanter maken van het model. Verder laat het hoofdstuk het belang zien van het differentiëren tussen verschillende typen interne stakeholders en toont het hoe interactie effect heeft op de rollen van ontwikkelaars, de betrokkenheid van actoren, en AI-ontwikkelingspraktijken binnen publieke organisaties.

Hoofdstuk 6 neemt als uitgangspunt dat AI van invloed zal zijn op professionals, of het nu gaat om zittende of nieuwe professionals. Er wordt verwacht dat specifieke kenmerken van professionals, zoals

autonomie, discretie en impliciete kennis, beïnvloed zullen worden door AI adoptie, bijvoorbeeld door discretie te verschuiven van zittende professionals naar nieuwe professionals of door de autonomie van zittende professionals in het veld te verminderen, vooral van degenen die beslissingen nemen. Daarom bestudeert hoofdstuk 6 middels een workshop en semigestructureerde interviews hoe professionals binnen inspecties verwachten dat generatieve AI (GenAI) hun werk zal beïnvloeden. Men verwacht dat GenAI bijdraagt aan veel repetitieve taken, maar men verwacht ook dat GenAI waarschijnlijk steeds meer impact zal hebben op meer complexe taken. Het hoofdstuk concludeert dat de opkomst van GenAI mogelijk leidt tot nog meer belang van stilzwijgende 'tacit' kennis, een sterke behoefte aan experimenteerruimte om leerprocessen van het werken met GenAI te faciliteren, en de rol van GenAI als sparringpartner in het dagelijkse werk van toezichthoudende professionals. GenAI als sparringpartner houdt in dat er een tweezijdige relatie ontstaan tussen professionals en GenAI, waarbij deze tools tevens tegenwicht kunnen bieden aan professionals. Dit vraagt om een sterk ontwikkelde professionaliteit.

De studies in dit proefschrift helpen samen om de hoofdvraag te beantwoorden: *Wat zijn de organisatorische implicaties van de adoptie van lerende algoritmen in publieke organisaties?* Logischerwijs is het antwoord op deze vraag meerzijdig. Het is echter duidelijk dat publieke AI adoptie grote implicaties heeft. Enerzijds toont deze dissertatie zien dat publieke AI adoptie resulteert in toenemende organisatorische complexiteit en incongruentie tussen hoe organisaties eruit zien, en wat er nodig is voor succesvolle AI adoptie. Fenomenen die dit illustreren zijn onder andere nieuwe grenzen (hoofdstuk 4), discretionaire afhankelijkheden tussen verschillende nieuwe en bestaande professionals (hoofdstuk 4 en 5), hoge tijdsinvesteringen zonder duidelijk rendement (hoofdstuk 4 en 5), en de noodzaak om nieuwe vaardigheden te verwerven (hoofdstuk 4, 5 en 6). Alhoewel dit niet verassend hoeft te zijn voor lezers die reeds bekend zijn met de effecten van nieuwe technologieën

op organisaties, lokken deze bevindingen wel vragen uit over of AI de investering waard is. AI wordt immers vaak genoemd als technologie die efficiëntie kan verhogen en dienstverlening kan verbeteren, maar de beschreven fenomenen lijken het tegenovergestelde te illustreren.

Maar deze conclusie laat slechts de helft van het verhaal zien. De andere helft vertelt een ander verhaal. Terwijl de buitenwereld AI steeds meer adopteert, kunnen publieke organisaties niet achterblijven. Terwijl zij AI adopteren, leren organisaties steeds beter wat er nodig is om AI te adopteren, ontstaan er nieuwe professionele rollen, en leren professionals als individu om te gaan met de dagelijkse realiteit van het werken met AI. De huidige incongruentie die in deze dissertatie is beschreven, impliceert daarom niet dat publieke organisaties AI niet moeten adopteren. Integendeel, deze dissertatie raadt aan dat publieke organisaties innoveren met AI.

Publieke organisaties, en de besluitvormers in deze organisaties, worden aangemoedigd om professionals te stimuleren die graag met AI willen werken, om te experimenteren met interactie, en om te begrijpen dat de adoptie van AI gepaard gaat met fouten, ambiguïteit en imperfectie. Alleen door te experimenteren kunnen deze publieke organisaties de lange termijn voordelen plukken die AI belooft. Deze organisaties zullen dat echter wel op bedachtzame wijze moeten doen, met oog voor de waarden die ze vertegenwoordigen, de strategische risico's van een te sterke afhankelijkheid van Amerikaanse GenAI oplossingen en de noodzaak van continue innovatie. Bedachtzaamheid houdt in dat publieke organisaties actief ontwikkelingen en signalen rondom AI zoeken en volgen. Het gaat erom voorzichtig te zijn en gelijktijdig AI adoptie te faciliteren, zonder passief te worden. In het AI-adoptie proces zullen deze organisaties en haar professionals worden geconfronteerd met allerlei fenomenen.

Fenomenen zoals de herverdeling van discretie van bestaande naar nieuwe professionals, de opkomst van nieuwe kennisgrenzen door

het waterbed effect, of een toenemende variëteit aan professionals en variëteit binnen professies. Ook zullen deze organisaties worden geconfronteerd met het blijven bestaan van hardnekkige grenzen die moeilijk te slechten blijken en de opkomst van hybride rollen, die een brug slaan tussen elkaar niet begrijpende beroepsgroepen. Dit is eigen aan het transformatieproces rondom AI adoptie en signaleert dat AI steeds meer een onderdeel van de organisatie wordt. Organisaties zullen echter niet het kind met het badwater moeten weggooien in dit moeilijke en ingewikkelde proces. Want in de wereld van AI is de professional nog altijd zeer belangrijk. Zonder goed ontwikkelde professionaliteit kan AI immers minder op waarde worden geschat en kan er geen tegenwicht worden geboden aan deze nieuwe technologie. Daarmee is het fundament van de AI-gedreven transformatie van publieke organisaties nog steeds afhankelijk van professionals van verschillende achtergronden en de interacties die zij hebben.

Sintesi

L'IA influenzerà significativamente le operazioni delle organizzazioni pubbliche. Ciò comporta molto più della semplice esecuzione più efficiente o efficace dei compiti esistenti di queste organizzazioni. L'adozione dell'IA[3] porterà a nuove possibilità per le organizzazioni pubbliche, difficili da prevedere al momento attuale. Di conseguenza, l'IA avrà anche importanti implicazioni organizzative. Il ruolo dei professionisti cambia, le strutture organizzative potrebbero richiedere adeguamenti e le dipendenze interne ed esterne tra i vari attori si modificano.

Con la crescente adozione dell'IA da parte delle organizzazioni pubbliche, è aumentato anche l'interesse scientifico per le implicazioni organizzative dell'adozione dell'IA. Sebbene la letteratura esistente riconosca sempre più che algoritmi e organizzazioni sono interconnessi, il concetto di "organizzazione" è spesso definito in modo minimale e vi è ancora scarsa attenzione alle componenti che costituiscono un'organizzazione. Inoltre, l'interrelazione tra organizzazione e tecnologia è stata indagata empiricamente solo in misura limitata.

Comprendere meglio l'interrelazione tra tecnologia e organizzazione, e prestare attenzione alle componenti che costituiscono l'organizzazione, non è rilevante solo da una prospettiva organizzativa. Esempi del passato, come l'algoritmo discriminatorio COMPAS utilizzato dai giudici negli Stati Uniti per valutare il rischio di recidiva, dimostrano che l'IA può avere conseguenze di vasta portata per i cittadini. Data l'interconnessione tra tecnologia e organizzazione, una migliore comprensione potrebbe potenzialmente ridurre anche l'impatto sui cittadini.

3 Per adozione dell'IA intendo non solo l'incorporazione dell'IA all'interno dei confini organizzativi, ma anche l'uso e lo sviluppo dell'IA nelle organizzazioni pubbliche. Nel fare ciò, impiego una definizione molto ampia di adozione dell'IA che trascende esplicitamente gli ambiti di molteplici aree di ricerca.

È necessaria una prospettiva più dettagliata sul concetto di «organizzazione». Un›organizzazione non è mai un fenomeno semplice, ma può essere analizzata attraverso diverse lenti. Questa dissertazione offre diverse di queste lenti e quindi una prospettiva più dettagliata sulle organizzazioni pubbliche che sviluppano e utilizzano l›IA. Se una prospettiva così di dettaglio non viene applicata, si rischia di fraintendere l'effetto che l'IA avrà sulle organizzazioni pubbliche e che le organizzazioni avranno sull'IA. Fraintendere questi effetti può portare a decisioni con impatti indesiderati nell'adozione dell'IA nel settore pubblico.

Questa dissertazione fornisce una prospettiva dettagliata sull'adozione dell'IA da parte delle organizzazioni pubbliche attraverso l'esame di tre livelli: l'individuo, l'organizzazione e il contesto. Per ciascun livello, questa dissertazione fornisce una specifica lente analitica.

- A livello individuale, l'attenzione è rivolta al singolo professionista. Questo professionista possiede autonomia e opera e prende decisioni sulla base dell'esperienza e della conoscenza tacita.
- Il livello organizzativo riguarda la molteplicità di attori coinvolti nell'adozione dell'IA – la collaborazione tra attori nuovi ed esistenti con background diversi e gli ostacoli che questo crea.
- L'analisi dell'ambiente organizzativo si fonda sulla prospettiva contingente applicata alle organizzazioni. Ciò implica un adattamento organizzativo ai fattori contestuali e ai requisiti che gli ambienti esterni impongono sulle organizzazioni pubbliche che adottano l'IA.

Attraverso questi livelli e le corrispondenti tre lenti, questa dissertazione risponde alla domanda: *Quali sono le implicazioni organizzative dell'adozione di algoritmi di apprendimento nelle organizzazioni pubbliche?*

Come evidenziato, la relazione tra IA e organizzazioni è intrinsecamente reciproca. Questa dissertazione risponde alla domanda di ricerca principale partendo dalla ricerca precedente. Tale ricerca evidenzia che gli algoritmi di IA trasformano le pratiche esistenti all'interno delle organizzazioni, mentre simultaneamente quelle pratiche esistenti all'interno dell'organizzazione possono influenzare il modo in cui gli algoritmi di IA vengono adottati. Questa dissertazione esamina la relazione bidirezionale tra algoritmi di IA e organizzazioni. Attraverso tre lenti chiave sulle organizzazioni, analizza come l'IA trasforma le organizzazioni mentre le pratiche esistenti influenzano simultaneamente l'adozione dell'IA.

Questa dissertazione si basa su tre distinti studi empirici esplorativi all'interno di organizzazioni pubbliche, un capitolo di libro di natura concettuale sui processi decisionali tramite l'IA nel settore pubblico e una riflessione sulla relazione tra IA generativa e la conoscenza tacita dei professionisti. Gli studi empirici corrispondono ai capitoli centrali di questa dissertazione, i capitoli 4, 5 e 6. Questi capitoli discutono le molteplici domande di ricerca che insieme rispondono alla domanda di ricerca principale. Tali domande di ricerca sono tutte correlate al processo bidirezionale in cui gli algoritmi di IA rimodellano i processi in cui sono incorporati e in cui le pratiche esistenti all'interno dell'organizzazione possono influenzare il modo in cui gli algoritmi di IA vengono adottati e utilizzati.

Il Capitolo 4 muove dall'idea che l'IA coinvolge una varietà di professionisti – esistenti e nuovi – che devono collaborare e interagire tra loro per sviluppare l'IA. Più specificamente, spesso sviluppatori e utenti finali creano insieme sistemi di IA. Tuttavia, la necessità di interazione è ostacolata da diverse sfide che possono limitarne l'efficacia, inclusi i confini tra sviluppatori e utenti finali. Il capitolo studia l'effetto che l'interazione ha sui confini tra sviluppatori e utenti finali nello sviluppo di ML attraverso uno studio di caso. Esistono confini di conoscenza riguardanti linguaggio, comprensione e obiettivi, e la domanda è cosa

produca l'interazione organizzata su questi confini. Il capitolo conclude che in superficie l'interazione contribuisce a sfumare i confini. Tuttavia, uno sguardo più approfondito rivela effetti più sfumati. L'interazione organizzata effettivamente sfuma i confini in una certa misura, ma confini profondamente radicati riguardanti linguaggio e comprensione persistono. Inoltre, emergono nuovi confini tra diversi tipi di utenti finali come conseguenza del coinvolgimento di solo alcuni utenti finali.

Il Capitolo 5 è correlato al fatto che le organizzazioni pubbliche affrontano diverse esigenze sociali nella loro adozione dell'IA. Ad esempio, che l'uso dell'IA da parte delle organizzazioni pubbliche debba essere spiegabile. Al pubblico, ma anche ai professionisti all'interno dell'organizzazione. Recentemente, l'interazione tra attori è stata associata come meccanismo per una maggiore spiegabilità. Data questa associazione, il Capitolo 5 esplora i benefici e le sfide dell'interazione sulla spiegabilità dei modelli di IA attraverso la ricerca-azione. Inoltre, esplora come questi benefici e sfide si manifestano. L'interazione è menzionata come condizione chiave per modelli di IA utili. La spiegabilità è spesso menzionata come proprietà di tali modelli. Il capitolo conclude che l'interazione tra gli attori coinvolti nell'IA aumenta la spiegabilità attraverso l'apprendimento organizzativo. In quanto tale, non contribuisce alla spiegabilità nel modo tecnico più tradizionale, aprendo il modello. Inoltre, il capitolo sottolinea l'importanza di differenziare tra diversi tipi di stakeholder interni e come le interazioni influenzano i ruoli degli sviluppatori, il coinvolgimento degli stakeholder e le pratiche di sviluppo dell'IA all'interno delle organizzazioni pubbliche.

Il Capitolo 6 parte dalla premessa che l'IA avrà un impatto sui professionisti, siano essi esistenti o nuovi. Si prevede che caratteristiche specifiche associate ai professionisti, come autonomia, discrezionalità e conoscenza tacita, saranno influenzate dall'adozione dell'IA, ad esempio spostando la discrezionalità dai professionisti esistenti ai nuovi professionisti o diminuendo l'autonomia dei professionisti esistenti sul campo, specialmente quelli che prendono decisioni. Pertanto, il Capitolo

6 studia come i professionisti all'interno degli ispettorati si aspettano che l'IA generativa influenzi il loro lavoro attraverso un workshop e interviste semi-strutturate. Si prevede che l'IA generativa contribuisca a molti compiti ripetitivi, ma sempre di più viene riconosciuto anche il probabile impatto dell'IA generativa su compiti più complessi. Questo capitolo conclude che l'emergere dell'IA generativa potrebbe portare a un'importanza ancora maggiore della conoscenza tacita, e a una forte necessità di spazio per la sperimentazione per facilitare i processi di apprendimento del lavoro con l'IA generativa, e il ruolo in particolare dell'IA generativa come interlocutore nel lavoro quotidiano dei professionisti della vigilanza. L'IA generativa come interlocutore significa che si sviluppa una relazione bidirezionale tra professionisti e IA generativa, dove questi strumenti possono anche sfidare e bilanciare le prospettive dei professionisti. Questo richiede una professionalità ben sviluppata.

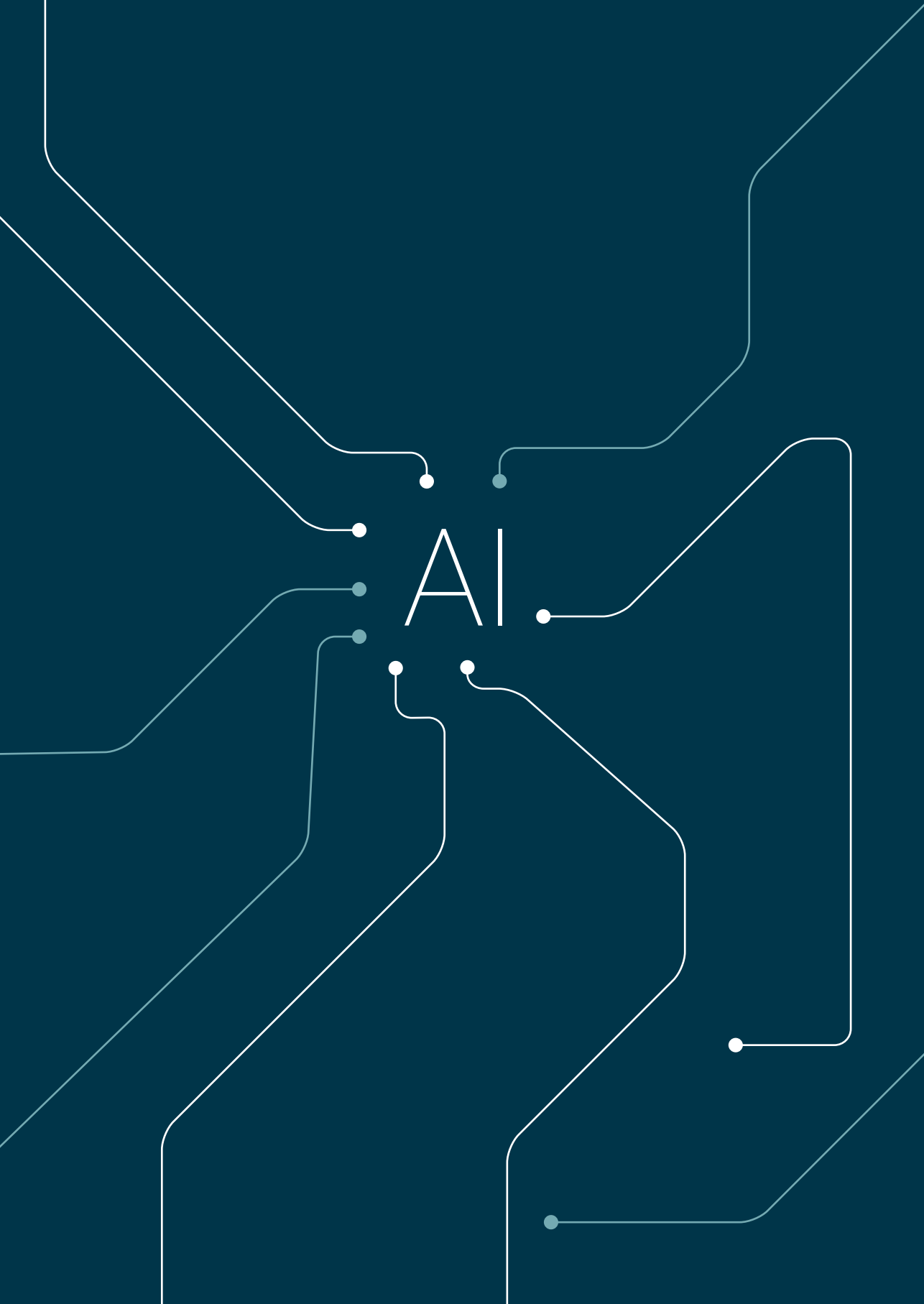
Gli studi raccolti in questa dissertazione contribuiscono collettivamente a rispondere alla domanda principale: *Quali sono le implicazioni organizzative dell'adozione di algoritmi di apprendimento nelle organizzazioni pubbliche?* Logicamente, la risposta a questa domanda è multiforme. È chiaro, tuttavia, che l'adozione dell'IA nel settore pubblico ha implicazioni significative. Da un lato, questa dissertazione mostra che l'adozione dell'IA nel settore pubblico comporta una maggiore complessità organizzativa e incongruenza tra come le organizzazioni sono strutturate e ciò che è necessario per un'adozione dell'IA di successo. Fenomeni che illustrano ciò includono nuovi confini (Capitolo 4), dipendenze discrezionali tra diversi professionisti (Capitoli 4 e 5), elevati investimenti di tempo senza ritorni chiari (Capitoli 4 e 5) e la necessità di acquisire nuove competenze (Capitoli 4, 5 e 6). Sebbene ciò possa non sorprendere i lettori già familiari con gli effetti delle nuove tecnologie sulle organizzazioni, questi risultati sollevano interrogativi sul fatto che l'IA valga l'investimento. Dopotutto, l'IA è spesso indicata come una tecnologia che può aumentare l'efficienza e migliorare l'erogazione dei servizi, ma i fenomeni descritti sembrano illustrare il contrario.

Tuttavia, questa conclusione racconta solo metà della storia. L'altra metà racconta una storia diversa. Man mano che il mondo esterno adotta sempre più l'IA, le organizzazioni pubbliche non possono permettersi di rimanere indietro. Mentre adottano l'IA, le organizzazioni apprendono ciò che è necessario per l'adozione dell'IA, emergono nuovi ruoli professionali e i singoli professionisti imparano a gestire la realtà quotidiana del lavoro con l'IA. L'attuale incongruenza descritta in questa dissertazione non implica quindi che le organizzazioni pubbliche non debbano adottare l'IA. Al contrario, questa dissertazione raccomanda che le organizzazioni pubbliche innovino con l'IA.

Le organizzazioni pubbliche, e i decisori al loro interno, sono incoraggiati a sostenere i professionisti desiderosi di lavorare con l'IA, a sperimentare con l'interazione e a comprendere che l'adozione dell'IA comporta errori, ambiguità e imperfezione. Solo attraverso la sperimentazione queste organizzazioni pubbliche possono raccogliere i benefici a lungo termine che l'IA promette. Queste organizzazioni devono, tuttavia, adottare l'IA in modo ponderato, considerando i valori che rappresentano, i rischi strategici di un'eccessiva dipendenza dalle soluzioni americane di IA generativa e la necessità di innovazione continua. Essere ponderati richiede che le organizzazioni pubbliche cerchino e monitorino attivamente gli sviluppi e i segnali relativi all'IA. Questo significa esercitare cautela mentre si facilita simultaneamente l'adozione dell'IA, rimanendo vigili piuttosto che passivi. Nel processo di adozione dell'IA, queste organizzazioni e i loro professionisti incontreranno vari fenomeni.

Fenomeni come la redistribuzione della discrezionalità dai professionisti esistenti a quelli nuovi, l'emergere di nuovi confini di conoscenza attraverso l'effetto vasi comunicanti, o una crescente varietà di professionisti e varietà all'interno delle professioni. Queste organizzazioni affronteranno anche confini persistenti che si dimostrano difficili da superare e l'emergere di ruoli ibridi che fungono da ponte tra gruppi professionali che non si comprendono reciprocamente. Questo è

intrinseco al processo di trasformazione che circonda l'adozione dell'IA e segnala che l'IA sta diventando sempre più parte dell'organizzazione. Tuttavia, le organizzazioni non dovrebbero gettare il bambino con l'acqua sporca in questo processo difficile e complesso. Perché nel mondo dell'IA, il professionista è ancora molto importante. Senza una professionalità ben sviluppata, l'IA non può essere adeguatamente valutata e non può essere offerto alcun contrappeso a questa nuova tecnologia. Pertanto, il fondamento della trasformazione delle organizzazioni pubbliche guidata dall'IA dipende ancora dai professionisti di varia formazione e dalle interazioni che essi hanno.


AI

1

Introduction

1.1. Why studying public AI matters

In the past years numerous examples have appeared where the use of self-learning algorithms within public organizations had severe negative effects. The COMPAS algorithm used by U.S. judges to assess recidivism risk (Angwin et al., 2016), and the UK A-level grading fiasco during COVID-19 (BBC, 2020), both showed algorithmic bias that disproportionately affected minorities. The Dutch Childcare Benefits Scandal, that impacted approximately 30.000 parents in the Netherlands, as they were victims of unjustified childcare benefit fraud suspicions by the Dutch Tax authorities further illustrates the risks of public AI systems (Algemene Rekenkamer, 2024).

Despite these well-documented incidents, AI adoption in government worldwide continues to grow. Particularly with the recent emergence of Generative AI (GenAI) applications, that provide yet another option next to the often-used machine learning (ML) models. The increasing development and use of these AI systems also applies to regulatory agencies – i.e. inspectorates - which hold a special position (Toezine, 2021). These organizations are responsible for large populations of inspectees while having limited resources to inspect and understand this population. To address this asymmetry, they increasingly turn to AI tools.

There are good reasons for the growing AI adoption. AI systems promise to increase the efficiency and accuracy of these public organizations. For instance, an inspector within the Human Environment and Transport Inspectorate might use AI to identify high-risk ships for inspection. The system can help to prioritize ships thereby making the inspectors decision-making process more efficient and informed. Additionally, the inspector may use a GenAI tool to assist with writing reports or other administrative tasks, thereby saving even more time. In doing so, AI gets

to the heart of what regulatory agencies do by helping to rationalize their selection processes.

Hence, we are at a key moment in the public sectors transformation. While some AI implementations have showed important challenges like discrimination, others highlight that the emerging AI adoption offers opportunities to enhance the efficiency and effectiveness of public organizations. Only by carefully studying public AI adoption can we make sure that these AI tools serve their intended purpose: improving public organizations, while maintaining accountability, fairness, and more. This dissertation contributes to this understanding by examining how public organizations, particularly regulatory agencies, are dealing with the complexities of AI development and use.

1.2. Why an organizational perspective on public AI is necessary

The literature on AI in public organizations has focused much on potential benefits and drawbacks. Scholars highlight AI's potential to increase efficiency and accuracy of public sector organizations (Mehr, 2017; Wirtz et al., 2019; Zuiderwijk et al., 2021). At the same time, these scholars also highlight drawbacks that AI might have for public organizations, including the risk of discrimination, privacy issues and a lack of transparency. However, as Lorenz (2024) argues, the approach of presenting benefits and drawbacks of AI algorithms as such may give the wrong impression. This approach often treats AI algorithms as stand-alone 'technical' objects and overlooks the organizations in which these algorithms are developed, adopted, and used.

The Dutch Childcare Benefits case illustrates why literature on AI and organizations must look beyond the technical focus to the organization. A report by the Dutch Data Protection Authority's revealed that end-

users of the risk classification model couldn't see which indicators led to specific risk scores. This opacity meant that professionals investigating flagged applications weren't aware of why these cases were deemed 'risky.' Had these professionals understood the reasoning, they might have alerted managers to potential issues, possibly preventing the scandal. And what about the top management of the organization. Why were they not aware of what was happening 'down below'? What was the role of AI? Did it amplify the distance between strategic leadership and operational professionals? This demonstrates that organizational choices influence AI's impact. It demonstrates that organizational choices—from responsibility allocation to communication channels—can have far-reaching consequences.

A more novel stream of public administration research builds on such insights and emphasizes that organizations matter when talking about AI algorithms (e.g. Meijer et al., 2021). These authors draw the perspective from the algorithms to the organizations and processes in which these AI algorithms are being adopted. Meijer, Lorenz, & Wessels (2021) show that AI algorithms reshape the processes in which they are embedded. They also show how existing practices within the organization may influence the way AI algorithms are adopted and used. For example, they found that differences in organizational norms between Dutch and German police organizations lead to different use of the AI systems in practice, potentially leading to different outcomes overall.

Although existing literature recognizes that algorithms and organizations are intertwined, the literature also offers opportunities for further research in this area. The relationship between AI algorithms and organizations can be described in more detail. Mostly because 'the organization' as a concept has not always been elaborated upon. Meaning that within the literature the concept 'the organization' is only minimally defined and the parts that make up the organization are not further clarified. Consequently, there is need for a more detailed and fine-grained understanding of what AI

means for organizations and vice-versa. If we accept that the organization matters, then we need to unpack the concept further. This thesis in part provides that elaboration.

This dissertation elaborates on the relationship between AI algorithms and organizations, and explores how AI algorithms transform organizational processes while organizational practices simultaneously shape AI adoption and use. This dissertation does this by taking three distinctive and detailed lenses on organizations. The next section will elaborate in more detail the lenses on the 'organization' that deserve additional scrutiny.

1.3. Three essential lenses on organizations and AI

"It's the people, stupid!"- unknown

I propose three complementary lenses on organizations. These lenses correspond to three intrinsically logical levels of analysis while looking at organizations: the individual level, the organizational level, and the contextual level. For each of these levels we have chosen a particular lens grounded in in existing literature. The lenses together offer a novel conceptual framework for understanding how AI systems shape - and are shaped by - public organizations. Table 1 provides an overview of the level of analysis and corresponding lenses. Within this dissertation the individual and organizational level receive the most attention. Because an organization cannot be separated from the environment it operates in, we also pay attention to the context.

Table 1: Overview of levels, lenses, relevance, and main concepts

Level	Lense	Relevance	Key concepts
Individual	Professional	AI impacts professional work	Autonomy, tacit knowledge, discretion
Organization	Multi-actor	AI adoption requires interaction between multiple actors	Knowledge boundaries, interaction, co-creation
Context	Contingency	Organizations do not use AI in a vacuum	Wickedness, accountability

For the individual level, I take the professional lens (de Bruijn, 2012; Lipsky, 1980; Mintzberg, 1979), focusing on the professionals' autonomy, discretion, and types of knowledge. The adoption of AI impacts professional work. It leads to the emergence of new types of professionals within public organizations. Also, AI undoubtedly impacts current incumbent professionals in how they carry out their work. AI is likely to complement more than just repetitive tasks, it may also take over more complex tasks. In doing so, it goes to the heart of professionalism. Hence, taking a professional lens is needed to understand how AI will impact public organizations.

For the organizational level, I take the multi-actor lens (De Bruijn & Ten Heuvelhof, 2018; van der Voort et al., 2019) . The professionals working on AI together consist of a group of multiple actors. The successful adoption of AI by public organizations requires much interaction between all these actors (Waardenburg & Huysman, 2022). Additionally, with the arrival of new AI-related professionals, such as data scientists, the organization becomes even more multi-actor. Following Mintzberg (1979) AI adoption may create tension between actors within the technostructure who promote AI-driven standardization and incumbent professionals that want to defend their discretion and autonomy. Hence, a multi-actor lens is needed to understand the impact of AI on public organizations.

Finally, we discuss the contextual level. Here I propose the contingency lens, which means that the success of the organization is partially determined by its fit with the context. Public organizations do not operate in a vacuum but exist in broader context. The context of public organizations is often wicked (Alford & Head, 2017). AI systems may be used to decide within that wicked context and its adoption is influenced by the demands of the context. Taking into account the wickedness of the organization is thus necessary to understand the impact of AI on organizations. Figure 1 depicts how the lenses relate to each other. All the lenses are more extensively elaborated in chapter 2.

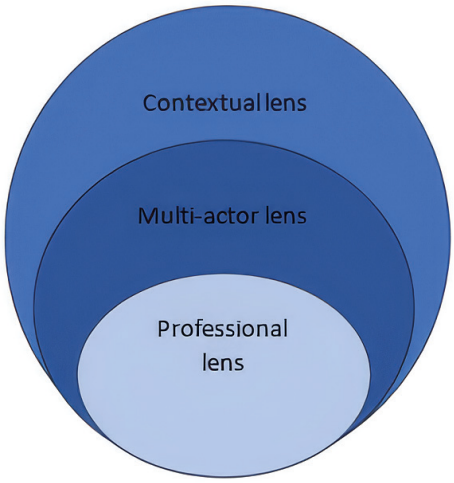


Figure 1: Relationship between lenses

1.3.1. Individual level: The professional lens

The professional lens holds that organizations consist of professionals. The literature offers several characteristics that can be attributed to professionals. Professionals are knowledge workers that use their knowledge to deal with complex tasks (Mintzberg, 1979). They operate relatively autonomously and have some form of discretion (Lipsky, 1980). The knowledge that they use to carry out complex tasks is often hard to standardize, sometimes even tacit (Howells, 1996) Therefore, the most

important form of control is often horizontal, meaning that professionals 'control' each other.

Within public organizations like regulatory agencies professionals exist in multiple places. A self-evident example is an inspector, who may autonomously decide to put an organization under restricted supervision or give a fine, depending on what that inspector finds appropriate. An inspector may be seen as incumbent professional, the type of professional who has been in the organization for a long time, and by nature had an important role. However, also the analysts providing information to these inspector departments may today be considered professionals. And what to think of legal experts, or people in the ICT department?

When AI enters into public organizations all these professionals may be impacted in many ways. First, AI systems may get in the way of the discretion of incumbent professionals, and against the wish of these professionals shift discretion partially to others or even to the AI system itself (Young et al., 2019). These others may be 'new professionals' like data scientists, that enter the organization because of the wish to deal with AI. Second, AI systems and their nature may be at tension with the type of knowledge that especially incumbent professionals use to carry out their tasks. AI can give the impression of clear and simple solutions, while reality – especially for professionals on the ground – is messier. Third, AI systems may decrease the autonomy of incumbent professionals on the ground, especially those making decisions.

The professional lens reveals how AI may fundamentally change the work of public sector professionals. From this lens focused on individual professionals, we must also consider that these professionals interact within the broader organizational context, which brings us to the multi-actor lens.

1.3.2. Organization level: The multi-actor lens

This lens recognizes that organizations and its surroundings consist of multiple actors. Organizations are not a unified entity but consist of various parts and departments which themselves consist of professionals. They on occasion work with professionals outside of the organization to accomplish their goals.

Public organizations frequently resemble what Mintzberg (1979) calls professional bureaucracies. Classical examples include hospitals, but also regulatory agencies are considered professional bureaucracies. Within these organizations professionals on the ground are the key actors. They often have autonomy to deal with complex tasks. Examples include doctors or, in the context of regulatory agencies, inspectors. A core trait of these organizations is variety. A variety of actors and departments exists. Mintzberg (1979) talks about variety and mentions the ongoing struggle between operators. i.e. cumbent professionals, with their wish for autonomy and the technostructure, who try to standardize the work of professionals. This tension partly explains why autonomy may complicate the necessary interaction between actors. These challenges also apply when public organizations dealing with complex issues start to adopt AI systems.

AI development is a process that inherently involves multiple actors and (Lorenz et al., 2022; van der Voort et al., 2019; Zweig et al., 2018) and leads to an increase in variety of actors. These actors - from data scientists and end-users to privacy specialists and project managers - all have a part to play at different points in the process by making consequential decisions, large and small. Figure 2 illustrates this line of reasoning. It brings forward that AI relies on the presence of a many actors, for example end-users, data scientists, privacy specialists, project managers, ICT specialists, architectural specialist and more.

The example highlights something else that may go without saying but is essential to mention. Actors need to *collaborate and interact with each other* in order to develop these AI systems (Lorenz et al., 2022; Waardenburg & Huysman, 2022). However, the need for Interaction is also where the shoe pinches, because interaction creates several coordination issues.

First, it may cause misalignment. AI asks for a chain of interactions between many – sometimes novel - actors scattered across the organization leading to a network of interactions. However, the structural configuration present in many public organizations struggles to accommodate the needed interaction. Second, there may be knowledge boundaries between different professional groups that hinder interaction. For example, developers and end-users need to interact, but also often seem to come from different worlds (Faraj et al., 2018; Lorenz et al., 2022; Waardenburg et al., 2018). Third, the needed interaction can be very demanding for particular actors – for example incumbent professionals - while the benefits are not always clear. Also, these demands do not always happen at either the desired moment from the perspective of the incumbent professionals or at the right time in the AI adoption process from the perspective of other actors. Hence, organizing the right interactions at the right time is not always possible.

The multi-actor lens shows how collaboration and interaction dynamics shape successful development and use of AI systems. However, public organizations do not exist in isolation. They are part of a broader societal context that impacts their goals and constraints. Additionally, organizational boundaries become more fluid because of increasing collaboration with external actors for successful AI adoption. Therefore, a contextual lens is crucial to understand AI's organizational impact.

To illustrate, consider the development of an algorithm for inspectors within a regulatory agency. Often this process starts with a data scientist, who is experimenting with AI algorithms and a data-set about a particular oversight area. At some point, this experimentation becomes more serious, and the data scientists has to make a choice about the type of algorithm that might be suitable for the particular situation. Additionally, in order for the system to be helpful, the data has to be labelled. Stakeholders that are able to label the data often need domain knowledge. Inspectors with domain knowledge, who are the intended end-users of the AI system, are therefore asked to label the data. To get their help, the data scientist approaches the inspection teams' manager. However, as the project moves from experimentation toward real-world application, a formal impact assessment becomes necessary. For this, the data scientist needs the help of a privacy specialist.

Figure 2: Illustration of multi-actor nature of AI algorithms

1.3.3. Contextual level: The contingency lens

The contextual lens means that public organizations do not exist as separate entities. They are part of an environment, in which they exist and operate. Based on contingency theory it follows that the environment partially dictates how these organizations are structured and act (Donaldson, 2001). The organization must be aligned with the environment to perform well. However, this complicates the story around professionals and the multi-actor organization even more.

First, for professionals it means that the environment places demand on their use of AI. For example, for some societal actors, AI use should be explainable (de Bruijn et al., 2021). Societal actors want to understand why and how AI is used by public organizations and what goes on behind the black box that AI is often called. However, the environment of public organizations is wicked (Head & Alford, 2015). The context of public organizations consists of multiple conflicting values and actors. Hence, it is impossible to satisfy all demands. A second implication has to deal with

actors within the environment. The necessary involvement of multiple actors is not limited to organizational boundaries. Public organizations often lack AI expertise (Delfos et al., 2022). Hence, they often rely on external organizations for successful AI adoption (Van Der Steen, 2024). What complicates both even more is the fact that professionals in public organizations use AI to decide about an environment in which there are also actors who use AI.

The contextual lens shows that context matters. The context dictates that there are multiple values at stake, multiple demands on public organizations, and a messy outside reality that organizations sometimes depend on but also have to look at. This has implications for the AI adoption. This justifies a closer look at AI within public organizations through the lens of the context.

1.4. Research questions and objectives

The interconnected lenses discussed above - professional, multi-actor, and contingency– show a complex puzzle about how AI systems transform public organizations. This puzzle is about the bidirectional relationship between AI algorithms and organizations. It is about how algorithms transform organizational processes while organizational practices simultaneously shape AI adoption and use.

To examine this puzzle, this dissertation investigates the adoption of AI algorithms in regulatory practice agencies, focusing particularly on the organizational implications. The overarching research question of this thesis is:

> *What are the organizational implications of the adoption of learning algorithms in public organizations?*

To answer this research question and different sub-questions, this thesis draws on concepts and ideas from various fields, including information science, organization science, and e-government. In answering these questions this dissertation departs from the premise that people together make the decisions that impact the technology, each other, and the rules and practices that are present within the organizations.

1.4.1. Research question 1

AI comes with a variety of professionals – incumbent and novel – that need to collaborate and interact with each other in order to develop the systems (Lorenz et al., 2022; Waardenburg & Huysman, 2022). A word sometimes used in this context is co-creation, mostly meaning that developers and end-users together create AI systems. However, the need to interact and co-create is impeded by several challenges that can limit its effectiveness, including boundaries between developers and end-users.

Research question one asks: *How does co-creation influence boundaries between developers and end-users in ML development and how do boundaries behave as a result?* The question connects to both the issue of interaction and that of professionalism. It focuses on the effects of co-creation i.e. interaction and collaboration, on boundaries between developers and end-users in the development of ML-tools. In doing so, it highlights how both types of actors, e.g. developers and end-users, have distinct types of knowledge and their own 'professional knowledge' that might get in the way of collaborating. As such, it highlights how interactive development of AI systems impacts organizations in practice. In that way it helps to answer the main RQ, because answering this sub-question teaches us about interactively developing AI systems, and what that might mean for the organization.

Overall, chapter 4 highlights how these professionals (developers and end-users) are impacted when they are facilitated to interact closely. The

chapter answers this question through a case study with ethnographic elements and draws from classical organization science theory (e.g. Bechky, 2003; Carlile, 2004) that discusses knowledge sharing between occupational communities and the difficulties that might arise.

1.4.2. Research question 2

Public organizations deal with various demands in their adoption of AI. For example, societal actors demands that AI use by public organizations should be explainable (de Bruijn et al., 2021; Miller, 2019). The AI systems should be explainable to the public, but also the professionals within the organization. Recently, interaction between actors has been associated as a mechanism for increased explainability (Bhatt et al., 2020; de Bruijn et al., 2021; Longo et al., 2024).

Research question two asks: *What are the benefits and challenges of an interactive and iterative development approach for explainability of AI models and how do benefits and challenges occur?* The question connects to both the underexplored areas of interaction and professionalism. The question is focused on exploring a collaborative development approach – characterized by interaction and iterations - and the benefits and challenges that it has. It is geared towards a particular feature of AI systems, namely explainability. The answers to this question will have implications for how organizations might organize themselves. As such, it connects to the main RQ by showing the effects of collaborative development on organizations.

Overall, chapter 5 will highlight how the interaction between stakeholders involved in AI increases explainability through organizational learning. It also highlights the importance of differentiating between different types internal stakeholders. Finally, it shows that the interactive process significantly impacts developer roles, stakeholder involvement, and AI development practices within public organizations.

1.4.3. Research question 3

AI will impact professionals in their daily work, whether incumbent or novel. Specific characteristics associated with professionals, like autonomy, discretion, and tacit knowledge are influenced with AI adoption, for example by shifting discretion from incumbent professionals to new professionals or to the AI system (Young et al., 2019), or by decreasing the autonomy of incumbent professionals on the ground, especially those making decisions.

Given the expected impact of AI, research question three asks: *How do professionals within inspectorates expect generative AI to affect their work?* The questions connects to the area of professionalism, since the question is about how professionals themselves look at the impact of GenAI on their work. It helps to answer the main research question, because this sub-questions is about the collaboration between professionals and GenAI. Through understanding how AI might impact the professionals work, we can understand how organizations might change to facilitate that change. The paper explores the main question in three steps. First it brings up a conceptualization of professionals working within regulatory agencies. Then it highlights empirical findings obtained by doing interviews and a workshop. Finally, it looks at these empirical findings and combines this with the lens of the earlier conceptualization of regulatory professionals.

Chapter 6 will highlight how the emergence of GenAI might lead to an even stronger importance of tacit knowledge (Howells, 1996), a strong need for room for experimentation to facilitate learning processes of working with GenAI, and the role of GenAI as a sparring partner in the daily work of oversight professionals. It highlights what organizations might do, to facilitate professionals working with GenAI.

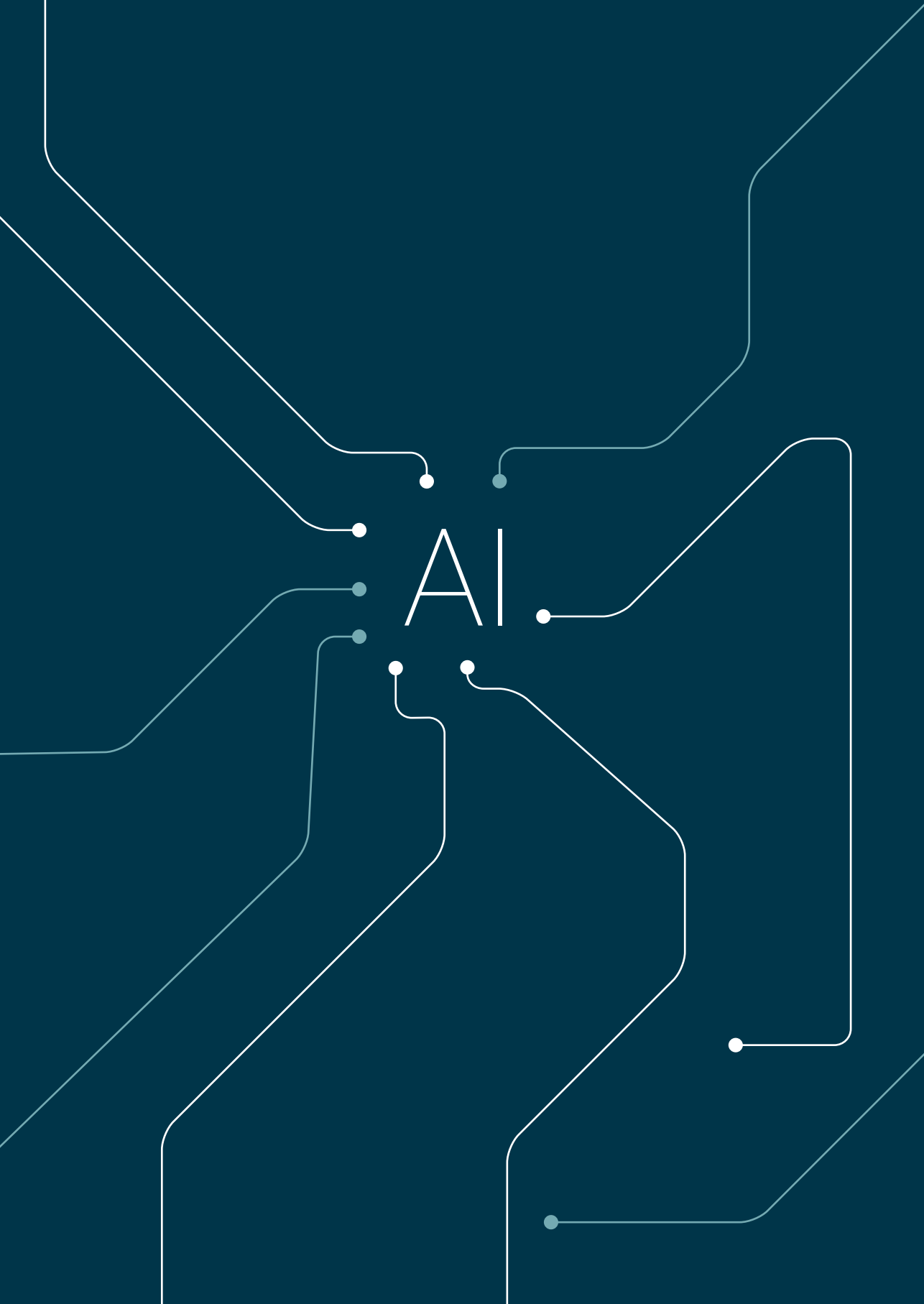
1.5. Dissertation outline

Table 2 highlights the outline of this thesis. Chapter two is an overarching theoretical chapter, that is partially based on an earlier published book chapter. Chapter three describes the methodology in a concise way. Since this dissertation is a collection of papers, each chapter that corresponds to journal articles describes the methodology separately. The chapters that correspond to articles all relate to a particular sub-question, that helps to answer the main questions.

Table 2: Thesis outline

Chapter	Title
Chapter 2	Key Theoretical Concepts
Chapter 3	Research Approach and Methodology
Chapter 4	Co-creation with Machine Learning: Towards a Dynamic Understanding of Knowledge Boundaries between Developers and End-users
Chapter 5	Explainable AI as a Learning Process: The Impact of an Interactive and Iterative Development Process on XAI for Organizational Stakeholders
Chapter 6	The Impact of Generative AI on Inspectorates: an Empirical Exploration among Professionals
Chapter 7	Discussion and Conclusion

1


AI

2

Key Theoretical Concepts

This chapter is partially based on the book chapter originally published as: van Krimpen, F. J., de Bruijn, J. A., & Arnaboldi, M. (2023). Machine Learning algorithms and public decision-making: A conceptual overview. In T. Rana, & L. Parker (Eds.), The Routledge Handbook of Public Sector Accounting (1st Edition ed.). Routledge - Taylor & Francis Group.

This chapter highlights the key theoretical concepts at the heart of this dissertation. First, artificial intelligence will be further elaborated. Then, starting with the professional lens, the three essential lenses on organizations and AI that are at the heart of this dissertation, are extensively discussed.

2.1. Artificial intelligence

2.1.1. What is it?

While AI is sometimes seen as a novel field of study, the field of AI and the study of machines that display 'intelligent behaviour' dates back many decades. Toosi et al. (2021) note that the term AI was coined in 1956 by John McCarthy during a workshop at Dartmouth College. However, even before then the study of artificial intelligence has received attention. For example, already in 1950 Alan Turing published an article in which he proposed a test to determine whether a task was carried out by a machine or a human. This test today is known as the 'Turing test' (Turing, 1950). These examples indicate that scholars have been thinking about 'intelligent machines' for a very long time. Hence, contrary to what some societal actors seem to believe, the field of AI is not new.

The AI field's long history has seen many different approaches and perspectives come and go over time. This is relevant, because contrasting and highlighting some of these approaches will help to understand one of the key challenges that exist with AI systems that are mostly used today. To understand todays challenges, we will draw our perspective towards artificial intelligence in general, expert systems, machine learning, and generative AI.

Artificial intelligence is often described as a field of study. However, this sometimes also makes it vague and it remains unclear what people mean when talking about AI. This necessitates a clearer definition of what is

meant when talking about AI. Artificial intelligence (AI) has been defined according to two abilities: 1) as the ability of machines to carry out tasks by displaying intelligent, human-like behaviour; and 2) as the ability of machines to behave rationally by perceiving the environment and taking actions to achieve some goals (Russel & Norvig, 2016). It is a domain in science that is concerned with the theory and practice of developing systems that exhibit characteristics we associate with intelligence in human behaviour (Tecuci, 2012). Forms that AI can take are, among others, case-based reasoning (CBR), cognitive mapping (CM), machine learning (ML) and artificial neural networks (ANN) (Sousa et al., 2019). However, such methods, especially as sophisticated as machine learning or artificial neural networks, is not what the field of AI started with.

2.1.2. Expert systems, Machine Learning and Generative AI

One of the earliest forms of AI are expert systems. Expert systems mimic the knowledge of an expert by combining a knowledge base with an inference engine (Gozalo-Brizuela & Garrido-Merchan, 2023). By applying if-else rules to the knowledge base, statements can be made about particular situations. For example, an expert system that mimics a doctor (an expert) can give the advice that someone has the flu, because conditions like having a headache, and having a fever, are being met. What makes it relevant to discuss these systems is the fact that such expert systems are inherently transparent, as the rules can be traced. Current forms of AI, and those studies within this thesis are not based on rule-based programming. Instead, they learn from data. The most well-known is machine learning. The rise of machine learning has been partially because of the increasing availability of large datasets and the increase in computational power.

Machine learning (ML) is seen as a specific area of artificial intelligence. (Mitchell, 1997) describes the field of ML as *''the study of computer algorithms that improve automatically through experience and by the use of data''* while (Samuel, 2000) defines it as *"a core branch of AI*

that aims to give computers the ability to learn without being explicitly programmed''. Thus, ML systems are systems that automatically learn from data. In essence, computers learn from data of various nature that is provided to them, which enables them to perform certain tasks (Alpaydin, 2020). Within machine learning, the used methods are supervised learning (most widely used), unsupervised learning, semi-supervised learning and reinforcement learning (Jordan & Mitchell, 2015). As supervised learn is often used, we draw specific attention to this method. More specifically, figure 3 shows how supervised learning can be used for classification.

Classification is a task in which the system takes an input, processes that input and assigns a 'class' value to that input (Domingos, 2012). An example of this application is the classification of emails into 'spam' or 'not spam' or the detection of credit-card fraud by labelling any given credit-card transaction 'fraud' or 'not fraud' (Brynjolfsson & Mitchell, 2017). Supervised machine learning might be applied as follows: In the 'fraud example' an algorithm has been trained on an existing dataset that contains cases with several 'independent' variables (such as income, current job and age) and a dependent variable 'fraud', that can be either 'yes' or 'no'. We call such a dataset 'labelled'. The algorithm learns from this set of examples and outputs a classifier (Domingos, 2012). That classifier can predict for future cases (based on the independent variables) whether that case is 'fraud' or not fraud'. A human expert acts as a teacher/supervisor (hence 'supervised learning'), since we show the computer the input and the correct answers with that particular input, and from that data, the computer itself learns the patterns. This can be contrasted to unsupervised learning, in which the computer is trained with unlabelled data.

Figure 3: Example of supervised Machine Learning

Recently, the field of AI, but also society in general, was shaken up with the emergence of ChatGPT. This publicly available Generative AI (GenAI) tool, based on a Large Language Model (LLM) has already seen many

upgrades and is now used by many people around the world. Generative AI represents a new paradigm in AI systems because these types of systems can mimic human creativity (Ritala et al., 2023) and can increase productivity and quality of knowledge work (Dell'Acqua et al., 2023). GenAI refers to a type of AI that can generate novel content (Gozalo-Brizuela & Garrido-Merchan, 2023). Similar to ML systems GenAI systems are trained on data. However, whereas ML systems are capable to 'discriminate' for example by classifying e-mail as spam or not spam, generative AI systems generate 'novel' content based on the enormous amount of data that it has been trained on. Generative AI systems can be increasingly useful for tasks such as writing business plans, legal documents, and software code. Furthermore, summarizing text, translating texts or writing drafts for e-mails or similar types of text are also easily done by GenAI tools. Some recent studies show that GenAI can outperform human workers on certain knowledge work tasks (Gilardi et al., 2023).

Table 3: Overview of types of AI relevant to this thesis

Type of AI	Description
Expert systems	Systems than give advice by combing a knowledge base with a rule-based inference engine.
Machine Learning	Systems that automatically learn from data and are able to 'discriminate' as a result.
Generative AI	Systems that learn from data and can generate novel content as a result.

2.1.3. Systems that learn from data

Looking at the evolution of AI systems over time an overarching finding emerges. Most current day AI applications, like machine learning and generative AI, are all about learning from data. *When I talk about AI or AI systems within this thesis, I refer to these type of systems that learn from data*. These types of systems, being based on ML or GenAI, are those being studied in this dissertation. Next to the use of AI, also the term 'learning algorithms' will occasionally be used. The fact that these

AI models learn from data has several consequences for organizations and its professionals, two of which deserve mentioning:

1. First, the key element of learning algorithms is at least some amount of opacity (Burrell, 2016). The opaqueness of these types of algorithms has consequences within organizations like regulatory agencies, as it is highly probable that the people asked to work with these systems will have at least a mild suspicion about the inner working of these systems.
2. Second, learning algorithms rely on the quality of the data. It is often said that '*the systems are as good as the data that goes into it*'. In organizations that deal with 'complex phenomena' like deciding who to inspect, the quality of the data is often dependent upon the input by domain professionals. Hence, for these types of learning algorithms, interaction between the data and the people who are potentially going to use the AI systems that are trained on that same data, is critical.

Both consequences will return in later chapters. Having defined AI within this dissertation, we now turn to the next step. The following sections describe the lenses highlighted in this thesis and the consequences that AI has for organizations, looking through these lens.

2.2. The professional lens and AI's impact

2.2.1. Why professionals are impacted by AI

The professional lens recognizes that many actors within public organizations may be considered professionals, meaning that they have some degree of autonomy and discretion, and rely on tacit knowledge (Howells, 1996) that is hard to standardize (de Bruijn, 2012; Lipsky, 1980; Mintzberg, 1979). These professionals do knowledge work and make choices based on their knowledge surrounding often complex topics.

Within regulatory agencies, the classical example would be the inspector. However, in modern day professional bureaucracies professionals exist in many places, as elaborated below.

The adoption of AI impacts the professional landscape in public organizations. Initially, AI was thought to mostly impact repetitive tasks (Viechnicki & Eggers, 2017), however a shift has occurred. It is now recognized that AI may also influence more complex tasks (Richthofen et al., 2022). Hence, whereas initially professionals were expected to be only moderately impacted by AI adoption because of the complexity of their work, there are reasons to believe that they may be impacted in a significant way.

This line of thinking is further illustrated by empirical examples. Van der Voort et al. (2019) show how algorithm adoption impacts decision-makers and data analysts roles, while Lebovitz et al. (2022), albeit focusing on doctors, show how AI may lead to varying levels of professional engagement among end-users. Both studies suggest that AI's impact on professionals may be significant and unpredictable.

In sum, scholars increasingly recognizes that AI may have a significant effect on the work of professionals. However, to fully understand how AI may impact professionals it is essential to incorporate the characteristics of professionals, such as autonomy, discretion, and their reliance on tacit knowledge.

2.2.2. How AI impacts professionals

As described above, the literature suggests that AI will impact professionals. However, how professionals may be impacted is mentioned less. This section describes several potential consequences of AI adoption for professionals and the professional bureaucracy. The consequences are:

1. Tension between AI-knowledge and tacit knowledge
2. The emergence of new professionals
3. Shifting discretion and autonomy

1. Tension between AI-knowledge and tacit knowledge[1]

A core trait of professionals is their reliance on tacit knowledge (de Bruijn, 2012; Howells, 1996). Tacit knowledge refers to knowledge and skills that individuals possess but are unable to express explicitly. It is often intuitive and nonverbal, acquired through personal experiences, observations, and practice over time (Polanyi, 1966). A common example is a medical professional who - without being able to explain it - sees that a patient suffers from sepsis based on some subtle clues. Maintaining this type of knowledge usually takes place in practice (Polanyi, 1966). The medical professionals maintains tacit knowledge trough seeing many patients. Relying on tacit knowledge is sometimes seen as the core of what makes someone a professional (de Bruijn, 2012), and it is exactly tacit knowledge that lives in tension with AI-based knowledge.

Most current-day AI models are based on data (Domingos, 2012). Based on training data AI models may give recommendations or generate content for a novel situation. Hence, AI models give the impression that based on that data clear-cut and unambiguous answers can be given in complex situations. With the growing belief that AI can not only contribute to repetitive tasks but also more complex tasks (Dell'Acqua et al., 2023; Fui-Hoon Nah et al., 2023), the question arises: what is the value of tacit knowledge? Hence, a tension between AI-knowledge and tacit knowledge reveals itself.

1 The arguments in this part are partly based on Van Krimpen, Van der Voort, De Bruijn. Generative AI and Professionals' Tacit Knowledge: Exploring Possible Relationships. (manuscript submitted to Public Administration Quarterly).

Several relationships between AI and tacit knowledge might occur:

- First, AI may *decrease the importance of tacit knowledge*. AI might be able to partially embody and make explicit the tacit knowledge of highly skilled and experienced workers. As a consequence, less skilled workers are able to use and leverage on this knowledge quicker than before (Brynjolfsson et al., 2023).
- Second, AI *may increase the importance of tacit knowledge*. As AI may embody the tacit knowledge of professionals (Brynjolfsson et al., 2023), the development of tacit knowledge is very important. Without tacit knowledge development, the resources for high quality AI systems may dry up. Hence, professionals should be given the opportunity to encounter complex situations that helps to develop their knowledge. Also, because research indicates that AI tool usage may lead to off-loading complex tasks to AI (Gerlich, 2025). Hence, the ability to use AI can be a hindrance to developing tacit knowledge.
- Third, *AI and tacit knowledge might serve as countervailing powers*. As tacit knowledge is often implicit, there may also be assumptions underlying it that are inconsistent with current reality. AI may act as a countervailing power, poking holes in the knowledge that professionals rely on. Vice-versa, professionals may counter the errors and oversimplifications of AI models. Professionals have the contextual knowledge and expertise to correct AI where it is wrong.

Figure 4: Continuum between over relying on AI versus tacit knowledge

2. The emergence of new professionals

AI adoption necessitates the introduction of new types of professionals, such as data scientists or privacy experts (Lorenz et al., 2022). These new types of professionals take their place along incumbent professionals like inspectors. Consequently, the variety in types of professionals is growing within the professional bureaucracies already full of many types of professionals. Waardenburg et al. (2022) give an empirical example, highlighting how within the police a novel professional role, being an 'intelligence' officer, was introduced because of the adoption of an AI system. That professional role acted as a liaison between the developers and end-users of the AI system.

The growing amount of new professionals leads to novel interfaces between professionals already present in the organization – incumbent professionals - and those entering. It heightens the complexity of the organization and its ability to function as a coherent whole. Furthermore, it leads to questions about power and importance. In the old organizations, the original professionals were the heroes, but who are the most important in the new organization? And what are the consequences for the incumbent professionals' autonomy?

3. Shifting discretion and limiting autonomy

Professionals rely on discretion and autonomy. Discretion allows them a degree of freedom to make decisions based on their judgment in situations faced with multiple options (Lipsky, 1980). For instance, an inspector choosing between restricted supervision or giving a fine, depending on what the inspector finds appropriate. Autonomy ensures that professionals make these decisions with relatively limited oversight.

The adoption of AI challenges both discretion and autonomy. First, AI systems can shift discretion from end-users to others like developers, or even to the system itself, creating 'artificial discretion' (Young et al., 2019). Shifting discretion to other actors happens when developers

make decisions about the type of algorithms or the parameters of the AI model. When they do this, they influence the choices that an end-user can make. Second, AI can decrease the autonomy of professionals. As noted, in an organization where norms are to follow recommendations of the AI system, the AI adoption significantly impacted what end-users could and could not do (Meijer et al., 2021). Hence AI adoption effectively limited their autonomy. Managers can decrease autonomy by providing instructions or setting norms that the AI system should be followed to a great extent.

2.3. The multi-actor lens and the challenges of interaction

2.3.1. Why multiple actors have to interact with each other

The multi-actor lens recognizes that organizations and their surroundings consist of multiple actors, who influence AI development and use in different ways. Additionally, it shows that these actors play crucial roles at different stages (van der Voort et al., 2019; Zweig et al., 2018). For example, some actors decide the type of algorithm, others decide if or how AI algorithms are implemented, while others label the training data.

However, that many actors are involved only tells halve the story. AI development requires more than multiple actors working independently, it demands interaction between these actors (Lorenz et al., 2022). Other scholars agree with this perspective and emphasize that AI development necessitates co-creation (Lebovitz et al., 2021), '*in which developers and users constantly share technical and work-related knowledge that requires them to constantly cross their boundaries*' (Waardenburg & Huysman, 2022). In other words, if one wants to develop AI algorithms, interaction between developers and end users to exchange knowledge is not an option, it is a necessity.

Interaction has also been described as a mechanism for increased explainability, which is seen as an important feature of learning algorithms, especially when used in critical contexts like regulatory agencies. While earlier XAI work focused on technical explainability approaches, recent research introduces how interaction between actors may increase explainability of AI models (de Bruijn et al., 2021; Longo et al., 2024; Meske et al., 2022).

In sum, both organization science and computer science literature increasingly endorse interaction as important for the development and use of successful and desired AI systems (Lebovitz et al., 2021; Longo et al., 2024). Yet, this is also where the crux lies, because interaction does not happen without struggle. This creates tension in AI development. On the one hand, interaction is needed, on the other hand, it comes with novel challenges that underscore that interaction is not the holy grail to realize successful AI systems.

2.3.2. The challenges of interaction

As described above interaction is an important condition for the development and use of helpful AI systems. However, that need for interaction within these public organizations is hindered because of several issues, shortly touched upon within the introduction. Below I unpack these issues in more detail and highlight the tensions that exist. The issues are:

1. Coordination issues
2. Knowledge boundaries
3. Demands and uncertain outcomes

1. Coordination issues

Many public organizations, including regulatory agencies, operate as professional bureaucracies. In professional bureaucracies coordination primarily occurs through standardisation of skills (Mintzberg, 1979). In

these organizations professionals tend to work in silos, while relying on specialized knowledge. The professionals work relatively autonomously within their silos and make decisions about their specific area. The organizational structure is designed to support the autonomous work of this diverse set of professionals.

However, the emergence of AI leads to a tension. AI adoption demands interaction between diverse professionals and interaction transversally through the organization. It demands that developers collaborate with end-users, that privacy specialists talk with developers, that project managers talk to end-users. This leads to an increasingly complex tangled web of necessary interactions, as novel interactions emerge, and the nature of interactions changes. For example, end-users, i.e. incumbent professionals, have to interact with novel professionals about data and AI. However, the structural configuration present in many public organizations struggles to accommodate the needed interaction. The 'old' structural configuration that facilitated the autonomy of professionals, hinders the novel interaction and coordination needed for successful AI adoption.

2. Knowledge boundaries

Professionals from different occupational communities often struggle to understand each other when they interact (Bechky, 2003; Carlile, 2004). This phenomenon is also present in AI development (Waardenburg et al., 2022). Especially between developers and end-users knowledge boundaries are present that can substantially impact both the development process and the eventual adoption of the system (Waardenburg & Huysman, 2022).

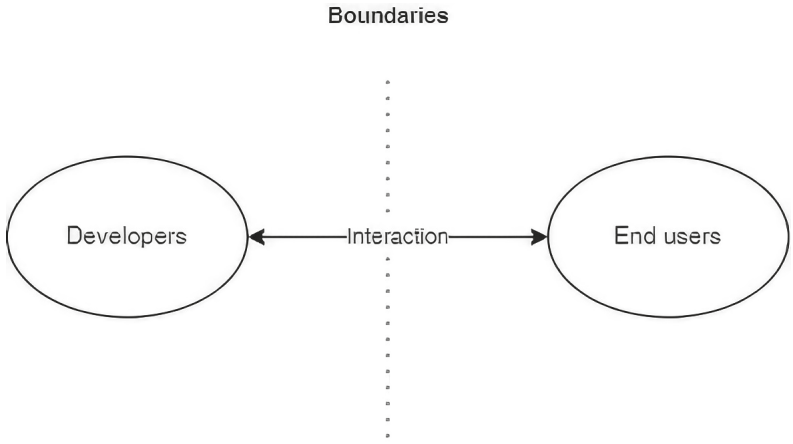


Figure 5: Boundaries between developers and end-users

Earlier work by van der Voort et al. (2019) empirically showed boundaries, highlighting that professional knowledge often competes with knowledge derived from 'data intelligence', for example obtained through AI systems. With this example they show that the occupational groups are from different worlds. Indeed, end-users and data professionals may interpret 'risk-oriented' work differently (van der Voort et al., 2021), highlighting their differences. Furthermore, crossing these boundaries comes with significant transaction costs. Are the efforts worth the reward? Beyond the studied link between developers and end-users, managers also play a crucial role because they facilitate the interactions that takes place, and managers also interact with developers and end-users.

In sum, there is a tension. On the one hand, interaction is essential to develop high quality AI systems. On the other hand, interaction exposes and potentially amplifies the differences between occupational communities. This begs the question of a better understanding of what interaction means in the development of AI systems.

3. Demands and uncertain outcomes

AI puts demands on the organization and its professionals, while the outcomes of these demands are uncertain. An important reason why

interactions in AI development are often needed is because end-users are asked to label data that is at the basis of the AI models (Lebovitz et al., 2021). The tacit knowledge that these incumbent professionals – the end-users - possess is often needed to give meaning to the data that the AI systems are trained on. Without their tacit knowledge, about the domain in which the AI model will be used, the AI might be useless.

At the same time, the fruits of their labour may be unclear to these intended end-users. These professionals must have confidence that AI investments will deliver meaningful returns. By contributing to labelling these professionals may even contribute to the (unwanted) transformation of their own job, for example leading to partial augmentation (Richthofen et al., 2022). Additionally, interaction comes with high transaction costs, as it takes time and effort away from professionals who are already heavily occupied. Both issues create a disincentive to contribute to the development of AI.

There is also a complicating temporal element to these demands. Finding the right time for interaction is hard. Involving end-users to early may place to much of a burden, while involving them too late may diminish their value and impact. Thus, there is a tension. Demands are placed on professionals, while the effects or perceived benefits of interaction and costs because of that interaction are unclear or even unwanted.

2.4. The contextual lens and the impact of the environment

2.4.1. Why the organizational context matters

The contextual lens recognizes that organizations exist within a broader environment, consisting of other organizations, citizens, and laws and norms that shape how they act. With the increasing adoption of AI by many organizations, the relationship between them become increasingly

complex. Public organizations are special in terms of the environment, because they are expected to realize 'public value' (Moore, 1995), which refers to the value created by public organization that benefit society.

The environment of public organizations can be defined as wicked (Head & Alford, 2015). It refers to an environment that consists of multiple conflicting values and problems for which there is no agreement about the problem itself. Prime examples of these problems are how to deal with poverty within a given society or how to organize the healthcare system. Both problems require balancing different values and approaches, while people disagree about these values.

As the environment partly dictates how these organizations are structured and act (Donaldson, 2001), organizations are expected to align with the environment. However, because of the environmental complexity and wickedness, there are various implications for professionals and the interactions between them, when taking into account the environment op the public organizations. They deal with an environment with conflicting requirements and values.

2.4.2. The challenges of the context

As described above, the literature points towards the importance of the context of organizations for AI adoption. The context of public organizations is special and leads to many issues that public organizations deal with when adopting AI. More specifically, it has implications for professionals, and for the multi-actor constellation they are. Here I unpack some of these issues in more detail. The first issue specifically links to the professional view, while the second issue links to the multi-actor view. The issues are:

1. The need for the context
2. The requirements of the context

1. The need for the context

Public organizations operate in an environment, where AI is more and more widespread, highlighted trough organizations that use AI tools to streamline their organizational processes or citizens that improve their daily activities. The growing AI adoption leads to two demands on organizations: they have to understand how stakeholders in the environment use AI, while also adopting AI themselves.

However, public organizations often lack AI expertise (Delfos et al., 2022). Hence, these organizations sometimes rely on external organizations for successful AI adoption (Van Der Steen, 2024). Such organizations include organizations offering themselves as advisors around AI adoption and implementation, suppliers of AI models and systems, or companies sharing AI-related knowledge. The dependency on external organizations leads to an interesting dynamic. Public organizations have to look at the environment for successful AI adoption, while they also need to provide the environment with AI driven services and understand what external actors are doing with AI.

2. The requirements of the context

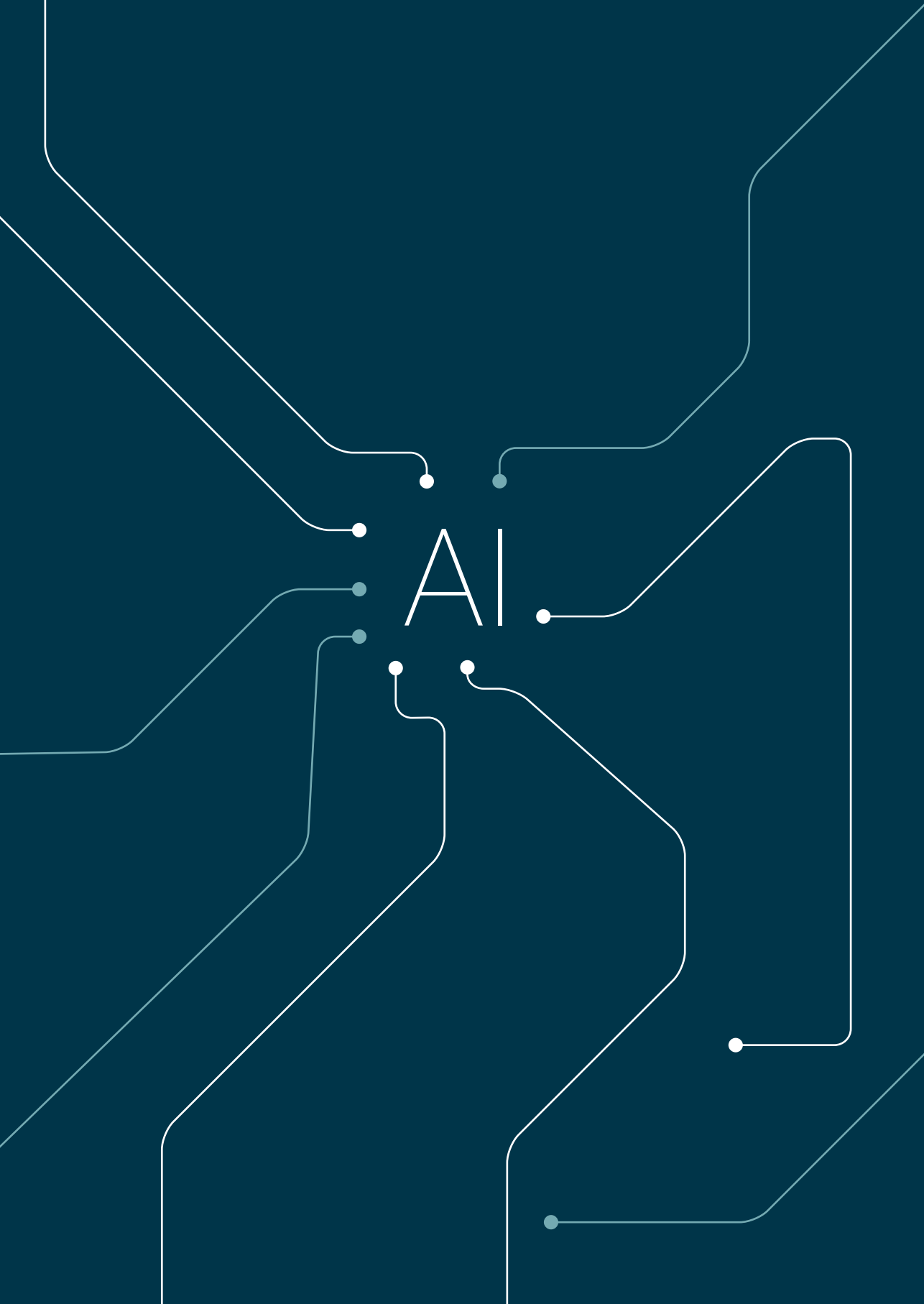
Public organizations operate in wicked environments, which – greatly simplified - means that stakeholders hold competing values (Alford & Head, 2017). Hence, public organizations, as organizations expected to create public value (Moore, 1995), must try to meet these conflicting demands as best they can. However, it is impossible to satisfy all the values. For example, some citizens want a transparent government, while others prefer tough enforcement over transparency.

The impossibility of meeting all requirements extends to AI adoption. Public organizations and professionals within can never meet all requirements, because the contextual actors asses the importance of the underlying values differently. Where some would find a particular level of AI transparency appropriate, others might disagree. The problem of

conflicting values is expected to be addressed by the public organization and professionals within.

This impossibility of meeting all requirements has consequences for organizing. It raises questions about whether public organizations understand and recognize these value conflicts, and how they cope with the challenges they present.

2


AI

3

Research Approach and Methodology

This section describes this dissertation overarching research approach and methodology. The main chapters – chapter 4, 5 and 6 - correspond to separate stand-alone papers, each with their own theoretical background, research questions and respective method to answer that questions. Therefore, this section only shortly discusses the research approach and methodology. I shortly discuss the research design, the main methods, the case selection, and the data collection. The data analysis is per paper is only discussed in detail in the chapters.

3.1. Research design

This dissertation relies on an exploratory and qualitative research design. Which is summarized in table 4.

Table 4: Research design

Research question	Methodological approach	Case
RQ1: How does co-creation influence boundaries between developers and end-users in ML development and how do boundaries behave as a result?	Exploratory case study	Human Environment and Transport Inspectorate (ILT)
RQ2: How can interactive and iterative development contribute to making AI explainable to organizational stakeholders?	Action research	'Check' (Italian public organization, pseudonym)
RQ3: How do professionals within inspectorates expect generative AI to affect their work?	Semi-structured interviews & Workshop	Health and Youth Care Inspectorate (IGJ) and Oversight Festival* *(additional data source)

The research employs an exploratory, qualitative design for two key reasons. First, the question of what the organizational implications of the interactive development and use of learning algorithms are involves

studying an emergent phenomenon. Second, understanding the complex effects that the development and use of AI in public organizations has demands an approach that allows for capturing nuanced perspectives and a deeper understanding (Hammarberg et al., 2016). Qualitative methods provide this flexibility and are about keeping the richness and complexity of the data. The next section highlights the main methods employed in the research design.

3.2. Research methods

The research design employs a mix of methods. I have carried out case study research, action research, semi-structed interviews, and a workshop. During the case study research and action research semi-structured interviews were the most important research method, but other techniques usually employed during case study research were also employed, e.g. document analysis and participant observation. The third research question has been researched through semi-structured interviews as the main method, in combination with a workshop.

All the chosen methods are strongly empirical and have an exploratory character. Through their nature the methods support the exploratory aim of this research. Through empirical insights this research aims to find effects of the development and use of learning algorithms. Due to the complexity of that aim and the fact that empirical research on this phenomenon is not yet matured, it is appropriate to use these type of methods. Below I will describe the main methods and how they have been applied during the research.

3.1.1. Case study

Case studies are often used to understand the dynamics of a real-world phenomenon that is not easily separated from its context. Case studies often deal with studying complex phenomena in which there will be

more variables of interest then data points (Yin, 2013). For such reasons, case studies often strongly rely on multiple sources of evidence. This so-called use of data triangulation for example happens trough collecting data via semi-structured interviews, observation and document analysis. Triangulation improves the validity of the outcome of the study (Yin, 2013).

I have used the case study method to study research question one. The phenomenon of the existence of boundaries between developers and end-users during ML development can be considered a complex real world phenomenon. Therefore, a case study is a suitable method to study this phenomenon. I have used data triangulation, and have used semi-structured interviews, observations and document analysis.

3.1.2. Semi-structured interviews

A semi-structured interview involves 'prepared questioning guided by identified themes in a consistent and systematic manner interposed with probes designed to elicit more elaborate responses' (Qu & Dumay, 2011). In essence, a semi-structured interview concerns doing an interview covering certain themes without a fixed list of questions. Furthermore, the interview is informal in tone and conversational. The interview allows for an open response in the participants' own words instead of a 'yes or no' type of answer (Longhurst, 2010). According to Barriball & While (1994), semi-structured interviews are well suited for the exploration of perceptions and opinions of respondents regarding complex issues.

The described characteristics highlight that semi-structured interviews are particularly useful for exploration around complex issues. Hence, I have used semi-structured interviews to help partially answer each of the research questions. During the interactive AI development processes in chapter two and three semi-structured interviews were the main method, supported by additional techniques used during case studies or action research. The study of research questions three was done with

a combination of interviews and workshop, making the semi-structured interviews an even more important method.

3.1.3. Action research

Action research is about combining theory and practice. It involves researches and practitioners and is focused on dealing with an actual problem situation in 'the real world' (Avison et al., 1999). Action research is usually carried out in an iterative process in which researchers and practitioners are acting together. By making changes in a real life problematic situation to see the effects of those changes practice can learn from theory, and theory can learn from practice. Thus, action research is a prime example of research that is done synergistically.

This type of research is well suited to understand processes in which researchers and practitioners work closely together and wherein the effects of novel interventions are tested. Hence, I have used action research to understand the benefits and challenges of an interactive and iterative development approach for explainability of AI models. Trough employing action research I could understand the effects that interaction to increase the quality of AI systems might have on the organization and the professionals within.

3.1.4. Workshops

Workshops as a research method are a relatively little studied phenomenon. We rely here primarily on work by Ørngreen and Levinsen (2017). Using a workshop as a research method usually aims to produce reliable and valid data. It often involves forward-looking processes, such as organizational change. Workshops are often used for studies that are unpredictable and related to practice. Ørngreen & Levinsen (2017) report on a study about E-learning and the use of video-conferencing in various organizations, in which they used workshops. Their example highlights that effects of implementing video-conferencing are unclear, and also very much related to a practical phenomenon. Workshops as a research

method often have inherent contradictions in them, as conflicting goals and roles may be present. Indeed, the workshop serves for the researchers to generate data, but at the same time it also prioritizes the needs of the participants and aims to give them something as well (Darsø, 2001).

3.2. Cases and case selection

The exploratory research has been carried out at two Dutch regulatory agencies – both inspectorates - at the national level and at a regulatory agency in Italy. Table 5 highlights the cases, their abbreviation, chapter, and the methodological approach. Semi-structured interviews at the Health and Youth Care inspectorate were combined with a workshop at the Dutch Oversight Festival 2024, where multiple regulatory agencies attended. Since the chapters contain a separate methodology, I will only describe here what regulatory agencies do, and why these organizations were selected as cases.

Table 5: Overview of cases

Case	Abbreviation	Chapter	Methodological approach
Human Environment and Transport Inspectorate	ILT	4	Exploratory case study
'Check' (Italian regulatory agency, pseudonym)	Check	5	Action research
Health and Youth Care Inspectorate and Oversight Festival	IGJ	6	Semi-structured interviews & workshop

3.2.1. What are regulatory agencies and why do they adopt AI?

Regulatory agencies are often relatively independent governmental organizations that deal with overseeing and enforcing laws, regulations,

and standards within specific sectors or industries (Koop, 2015). A specific form of regulatory agencies are inspectorates. Even more then 'normal' regulatory agencies, these inspectorates are primarily focused on inspection, monitoring, and enforcement of compliance with laws, regulations, and standards. Despite the fact that the distribution between what is a regulatory agency and what an inspectorate is not necessarily clear-cut, this provides an indication of the types of organizations within the scope of this thesis. These organizations do their work through research – including physical inspections - and inspection reports, annual reports and publications. Through this these inspectorates try to build trust. They do this by acting as a counterforce within the government and by pointing out what is going wrong as well as what is going well.[1]

Regulatory agencies often deal with a large number of potential inspectees while they have limited resources. The organizations are therefore highly reliant on risk-driven regulatory approaches (Lorenz, 2024), which loosely refers to using systems to identify inspectees that are at high risk of non-compliance (Hutter, 2005). Hence, the rationale for risk based regulation is that the regulatory agencies should visit those organizations where risks are highest.

Here AI comes into the picture, as the hoped-for benefits of AI systems, efficiency and accuracy, are evident in regulatory agencies desire to use AI for risk-based regulation (van der Voort et al., 2021; Veale & Brass, 2019). The idea is that an AI system – through its data driven nature - can help to determine the 'riskiest' organizations, for example through prioritization based on risk-score (Zuiderwijk et al., 2021). Using AI systems for these purposes goes to the heart of what regulatory agencies do, as AI holds the promise of making the selection process more accurate or efficient.

1 https://www.rijksinspecties.nl/over-de-inspectieraad/over-de-rijksinspecties

However, with the emergence of GenAI the possibilities for regulatory agencies have become even greater. As noted in chapter 2, GenAI can be useful for various tasks that regulatory professionals engage in, such as writing reports, summarizing text, or writing drafts for similar types of text (Dell'Acqua et al., 2023). Hence, also GenAI systems may help to increase efficiency and accuracy of regulatory agencies. However, recent work even suggest that GenAI may touch the core or professional work, as it may be able to partially embody tacit knowledge (Brynjolfsson et al., 2023), thereby again touching the core of what regulatory agencies do.

3.2.2. Why does this dissertation study regulatory agencies?

I studied regulatory agencies for the following reasons: 1) High probability of engaging with AI; 2) Most likely to be open to inquiry by researchers; 3) Contrasting nature of traditional way of working with nature of AI.

First, regulatory agencies were expected to have a high probably of engaging with AI. As noted, regulatory agencies often deal with a large number of potential inspectees while they have limited resources. Given this discrepancy we expected the regulatory agencies to be open to the potential of technologies such as AI. After all, these offer the potential for easier control of the large group of inspectees. Second, given their naturally outward-looking focus, we expected that regulatory agencies were probably open to inquiry by researchers. Last, we saw the contrasting nature of the traditional way of working of many regulatory agencies with the nature of AI, as a given that could potentially lead to interesting insights. Traditionally, within regulatory agencies, inspectors or similar roles are the focal actors. They do 'the most important' work. These actors make important choices, which they often base on all kinds of knowledge gained over the years, including tacit knowledge. AI's nature is different. It is sometimes said that the consequence of AI is going to be that such professionals are less important. This contrast ensures that right in these types of organizations, it is interesting to study the effect of AI adoption.

3.4. Data collection

For a detailed description of data collection per paper please refer to the method section of chapter 4, 5, and 6. Here I only aim to provide an overview of the data collection as a whole.

At all the organizations studied, access had to be obtained first. At both Dutch inspectorates I got in touch with a contact person at the relevant inspections, through the co-authors of the papers. After this initial contact, I coordinated contacts with these organizations. At the Italian regulatory agency, I was invited to be part of a research team. Here, the contact when mostly through one of the co-authors. The oversight festival happened at a set date. In order to be able to present here, we submitted a presentation proposal, and were admitted based on that proposal.

Table 6: Data collection overview

Case	When?	What?
Human Environment and Transport Inspectorate	March 2022 – February 2023	14 semi-structured interviews, 25 hours presence at meetings, 13 documents
'Check' (Italian regulatory agency, pseudonym)	Sep 2022 – Nov 2023	8 semi-structured (group) interviews; attendance/ participation at 9 meetings, 3 documents
Health and Youth Care Inspectorate and Oversight Festival	May 2024 – July 2024; Oversight Festival: June 18th 2024	7 semi-structured interviews ;1 hour workshop with 60 attendees leading to 16 documents

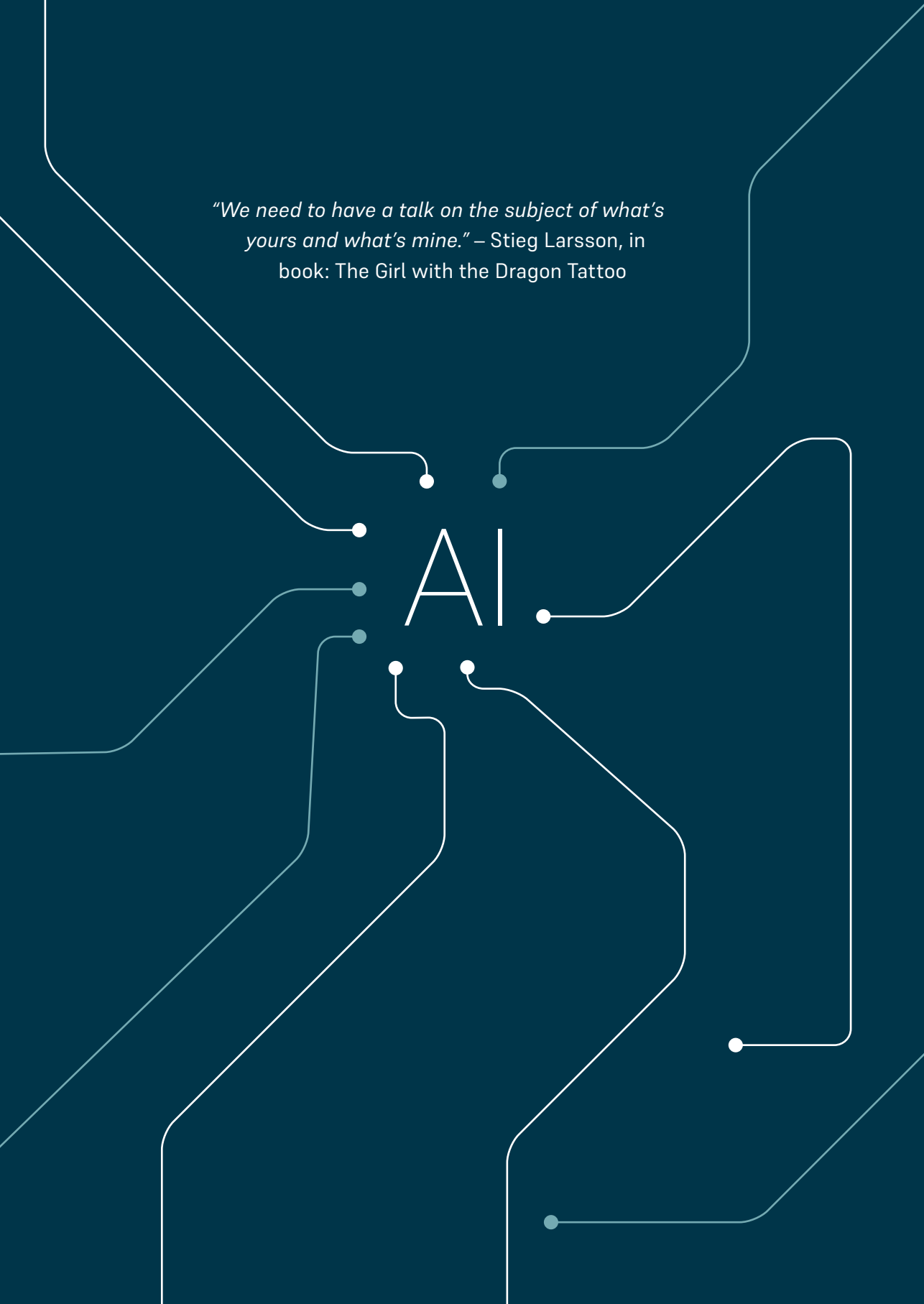
For both Dutch cases, the process leading up to data collection consisted of multiple informal talks. Within these formal talks we discussed the interests of our research and the pressing questions present within these

organizations. For example, at the Human Environment and Transport Inspectorate, while talking with innovation-minded professionals, it became clear that they had considerable difficulty with implementing innovations in practice. This sparked interest in both to study reasons for these difficulties more closely. Similarly, at the Health and Youth Care Inspectorate the swift emergence of GenAI led to questions about the impact on the organization. Since the co-authors and me also had interest in the impact of GenAI on public professionals, a research proposal could be made. At Check, the general direction of the project was negotiated by more senior co-authors.

An overview of the data collected per case is in Table 6. The data collection started in March 2022 with an online meeting at ILT's data science team. Here I could meet the team members, introduce my research, and listen in on the meeting. Presence at this meeting and a follow-up e-mail about my research helped me to establish relationships with people in the organization and with planning interviews. Throughout the data collection I attended both physical and digital meetings. Towards the end of data collection at the ILT, data collection at Check started. Here, all data were collected virtually. Before official data collection a plenary meeting was organized that I could digitally attend. In May 2024 data collection started at the IGJ. All interviews were conducted online. The Oversight Festival workshop was an in-person event, where we hosted participants for approximately 60 minutes. For each of the cases, data analysis started during the data collection. For more information about data collection, please refer to chapter 4, 5, and 6.

3

"We need to have a talk on the subject of what's yours and what's mine." – Stieg Larsson, in book: The Girl with the Dragon Tattoo



4

Co-creation with Machine Learning: Towards a Dynamic Understanding of Knowledge Boundaries between Developers and End-users

This chapter is originally published as a journal paper: van Krimpen, F., & van der Voort, H. (2025). Co-creation with machine learning: Towards a dynamic understanding of knowledge boundaries between developers and end-users. Information and Organization, 35(2). https://doi.org/10.1016/j.infoandorg.2025.100574

Abstract

The impact of machine learning within public organizations relies on coordinated effort over the functional chain from data generation to decision-making. This coordination faces challenges due to the separation between data intelligence departments and operational intelligence. Through theory about knowledge sharing between occupational communities and a case study at a Dutch inspectorate, we explore knowledge boundaries between machine learning developers and end-users and the effects of co-creation. Our analysis reveals that knowledge boundaries are dynamic, with boundaries blurring, persisting, and emerging under the influence of co-creation. Especially the emergence of boundaries is surprising and suggests the presence of a waterbed effect. Furthermore, knowledge boundaries are layered phenomena, with some boundary types more prone to change than others. Understanding knowledge boundaries and their dynamics better can be crucial for improving the intended impact of ML for organizations.

Keywords
Machine learning, coordination, knowledge boundaries, occupational communities, co-creation, knowledge sharing, waterbed effects

4.1. Introduction

Machine learning algorithms are increasingly deployed in public sector organizations, including to support critical decision-making processes. For instance, ML algorithms are now being used to inform complex decisions such as evaluating early prison eligibility (Berk, 2017) or the allocation of enforcement resources within regulatory agencies (Yeung, 2018). The use of algorithms to facilitate regulatory tasks is called 'algorithmic regulation.' Algorithms are increasingly used to coordinate the behaviour of inspectees or manage risks within a particular area of inspectees (Ulbricht & Yeung, 2022). These public organizations introduce ML algorithms to enhance decision-making accuracy (Domingos, 2012; Eggers et al., 2017) or streamline data analysis and interpretation (Alexopoulos et al., 2019).

However, ML development presents many challenges, including transparency (Burrell, 2016), accountability (Veale et al., 2018), and discrimination (Barocas & Selbst, 2016). Recent scholarly attention has increasingly highlighted organizations as the context in which both promises and challenges of ML materialize (D. Bailey et al., 2019; Faraj et al., 2018; Richthofen et al., 2022; Waardenburg & Huysman, 2022). An emerging challenge is specialization. Specialization is an issue when different specialized actors struggle to collaborate effectively in ML development due to divergent types of knowledge. Machine learning development requires specific knowledge and skills, typically beyond the expertise of end-users or decision-makers (Richthofen et al., 2022; Waardenburg et al., 2018). Consequently, organizations frequently develop ML-based innovative techniques through dedicated sub-units (H. G. van der Voort et al., 2019). Lorenz et al. (2022) argue that this separation may affect the effective development and use of ML algorithms.

Similarly, Waardenburg and Huysman (2022) theorize that blurring boundaries between end-users and developers seems needed to create helpful self-learning ML systems. However, these scholars also acknowledge that in practice boundaries often persist, even when interventions such as co-creating ML are taken to blur these boundaries (Waardenburg et al., 2022). A key concept is "co-creation". It refers to constant knowledge sharing between developers and end-users to develop ML systems (Lebovitz et al., 2021; Lorenz et al., 2022; Waardenburg & Huysman, 2022). Despite a co-creation intervention, boundaries often persist. This persistence necessitates a nuanced and deep-rooted understanding of these boundaries. Particularly, how the boundaries respond to interventions.

Investigating boundaries and their behaviour is important for multiple reasons. First, boundaries may significantly impact ML tool performance in public organizations. Second, boundaries may also lead to important ethical trade-offs falling between the cracks of different occupational communities, like developers and end-users. Lastly, understanding the effects of co-creation on boundaries in ML development is important because not knowing the effects could lead to unintended effects occurring. Ulbricht and Yeung (2022) argue that studying algorithms and their development in their respective contexts is essential for comprehensive insight into their effects.

This paper examines the dynamics of boundaries between machine learning developers and end-users. It focuses on a case where boundaries were intentionally removed to facilitate interaction. Within this case, we explored how boundaries both blur, persist and emerge. This informs us about the boundaries and their behaviour and cooperation between developers and end-users. Our main question is: *How does co-creation influence boundaries between developers and end-users in ML development and how do boundaries behave as a result?*

In our theoretical section, we will use knowledge management theory to explore the boundaries, most notably theories about knowledge boundaries (Carlile, 2004) and knowledge sharing between occupational communities (Bechky, 2003). We further explore the dynamics of boundaries following mostly a prior study by Waardenburg & Huysman (2022). In section 3 we introduce a case study at the Dutch Human Environment and Transport Inspectorate (ILT). This inspection agency introduced a program management approach to intentionally blur the boundaries between developers and end-users of ML. The findings are presented in section 4. After that, section 5 focuses on the discussion of the implications of these findings and links them to the theory. Furthermore, it discusses the limitations of the study and potential future research. Finally, section 6 concludes.

The paper's theoretical contributions are twofold. Firstly, the paper enriches the organizational literature that discusses knowledge sharing and boundaries by adding new and specific insights about knowledge boundaries in the context of the development and use of ML. Secondly, the paper adds to the organizational literature by exploring the mechanisms through which new ways of co-creation influences the behaviour of boundaries and how they blur, persist, or emerge in the development of public ML algorithms.

4.2. Knowledge Boundaries between Machine Learning developers and end-users

4.2.1. Blurring boundaries through the co-creation of ML systems

The increasing presence of ML systems in organizational settings has spurred the attention of organizational scholars who study the consequences for work and organizing (D. Bailey et al., 2019; van den

Broek et al., 2021). More specifically, the study of knowledge boundaries related to ML development has received increasing interest (e.g. Faraj et al., 2018; Waardenburg & Huysman, 2022).

Some studies have focused their attention on how to deal with boundaries. To address boundaries requires collaboration. Mayer et.al (2023) propose managers guidelines to promote collaborative AI development. They include to create a shared vision, to build a collective understanding, and to develop complementary abilities. Programme management (Ferns, 1991; Pellegrinelli, 1997) offers a similar intervention, facilitating cross-departmental collaboration through simultaneous multi-project coordination.

Other scholars studied boundaries and strategies to overcome these even in more detail. Waardenburg et. al (2022) argue that knowledge sharing in ML development occurs through sharing practices. Participating in shared practices would help actors develop a shared understanding (Brown & Duguid, 1991; Lave & Wenger, 1991; Orr, 1996). Hence, Waardenburg and Huysman (2022) advocate for a co-creation perspective in the case of the development, implementation, and use of self-learning ML algorithms. They argue for the sharing of practices or the emergence of new communities of practice. Their work highlights that '*new fields of practice can emerge that allow for knowledge to be shared between communities*' (Levina & Vaast, 2005).

Co-creation suggests that overcoming boundaries requires to move from a perspective of co-existence of stakeholders in the development of ML systems towards a view that sees stakeholders as co-creating the systems (Holmström & Hällgren, 2022; van den Broek et al., 2021; Waardenburg & Huysman, 2022). Co-existence implies hardly any interaction between stakeholders, while co-creation implies the opposite. Such a practice-based perspective on knowledge sharing (Orlikowski, 2002) emphasizes that '*developers and users constantly share technical*

and work-related knowledge that requires them to constantly cross their boundaries' (Waardenburg & Huysman, 2022, p.2).

However, empirical evidence suggests that crossing boundaries in ML development remains challenging, because boundaries persist (Waardenburg & Huysman, 2022). Waardenburg et al. (2022) explored interventions, such as deploying intelligence officers as brokers between developers and end-users. The broker role they studied was meant to resolve a semantic boundary between the communities and can be seen as an intervention to blur boundaries. Paradoxically, this intervention to blur boundaries instead had unintended consequences. The algorithmic broker role had the effect of the emergence of novel boundaries.

The example illustrates a critical insight: organizational interventions aimed at blurring boundaries can unexpectedly reinforce or create novel boundaries elsewhere. The example illustrates that despite deliberate efforts boundaries can persist. This raises the question of an in-depth understanding of what these persisting boundaries are. To do this, there is a need to understand more about the literature on knowledge boundaries.

4.2.2 Knowledge boundaries in ML development

Organization science and knowledge management literature has explored knowledge boundaries and ways to overcome them (Bechky, 2003; Carlile, 2002, 2004; Orlikowski, 2002). Carlile (2004) discusses an integrative framework for managing knowledge across boundaries, delineating three distinct boundary types.

The first is the syntactic boundary, informed by an information-processing perspective on boundaries (Galbraith, 1973; Lawrence & Lorsch, 1967). To overcome a syntactic boundary knowledge can simply be transferred across it. According to Carlile (2004), a second boundary type is semantic. Semantic boundaries arise from the *'differences in meanings, assumptions, and contexts'* (Kellogg et al., 2006) that create ambiguity

between organizational actors. Lastly, Carlile (2004) discusses the pragmatic boundary. The pragmatic boundary reveals the political aspects of knowledge and shows that knowledge is at stake (Kellogg et al., 2006). This perspective acknowledges that organizational stakeholders are invested in their knowledge and tie their measure of worth to it (Carlile, 2004).

Carlile's (2004) typology acknowledges the diversity of knowledge boundaries and delivers a conceptual framework for scholars to respect that diversity. Carlile (2004) also seems to imply that knowledge is manageable and transferable across boundaries, which makes knowledge resemble a tangible resource. Though the typology of boundaries clarifies their diversity, an emphasis on the boundaries may move the attention away from organizational stakeholders being bounded.

Bechky (2003) shifts the focus from the boundaries to the occupational communities these boundaries delineate. An occupational community refers to a community of organizational stakeholders that are part of the same occupation (Van Maanen & Barley, 1984). According to Bechky (2003), knowledge does not exist as a static entity within such a community. Instead, knowledge develops in relation to the activities that people engage in. Knowledge is constructed and situated in the organizational communities in which it is present (Van Maanen & Barley, 1984; Weick, 1979). It is argued that knowledge is embedded in practices (Brown & Duguid, 1991). This means that knowledge cannot be seen as a tangible resource that can be transferred from one community to the other. Instead, knowledge is embedded in daily procedures and routines that occupational communities engage in (Nonaka & von Krogh, 2009).

Given this view, Bechky (2003) argues that overcoming boundaries requires interaction, direct communication, and physical proximity. She highlights that knowledge sharing depends on social interactions, shared experiences and common contexts between professional communities.

Bechky' s (2003) view reveals that knowledge sharing is dynamic. This perspective aligns with Orlikowski's (2002) conception of 'knowing' as an ongoing social accomplishment. As a result, knowledge is shaped by everyday practices and as actors engage the world in practice.

While Bechky (2003) and Orlikowski (2002) highlight the dynamic nature of knowledge and knowledge sharing, existing literature has difficulty to translate these findings to understanding knowledge boundaries. Despite literature on communities emphasizing dynamics, the accounts on boundaries seem more static. Yet from the starting point that knowledge boundaries are diverse, it could easily be theorized that different types of boundaries may also exist at the same time, and that these boundaries may be subject to change. It is our assumption that an account on the dynamics of knowledge boundaries would add to our understanding of the adoption of machine learning in organizations.

This study aims to develop a dynamic account on knowledge boundaries in the context of machine learning, focusing on a conscious attempt to share knowledge across boundaries while developing ML (i.e. co-creation through program management). This study pursues two contributions. First, we want to highlight the complexity of co-creating ML and study the dynamics of the boundaries during this attempt. We answer to the call to do more empirical research in the development and use of ML and the boundaries that might be present (Lorenz et al., 2022; Waardenburg & Huysman, 2022). Second, we aim to contribute to boundaries literature (Bechky, 2003; Carlile, 2004; Orlikowski, 2002) and highlight the connectedness and co-existence of boundaries. In the study we will focus on both the boundaries and practices by occupational communities.

4.3. Methods

This study employs an exploratory, inductive, single case study to develop existing theory (Siggelkow, 2007). This method enables to further expand on previous theories about overcoming knowledge boundaries and to bring a rich empirical picture to the research of ML technologies in public organizations.

4.3.1 Case description

The *Inspectie voor Leefomgeving en Transport* (The Human Environment and Transport Inspectorate; ILT from now on) implemented the "Less Greenhouse Gases" programme from begin 2018 to the end of 2022, aimed at reducing emissions of greenhouse gases (F-gases) and ozone-depleting substances (OAS). The ILT started the programme to target the resources of the ILT towards topics where societal risks are the biggest and the greatest social damage can occur.

Programmes are '*a temporary way of working together, aimed at pursuing certain goals, which contribute to achieving the strategy of the organisation(s)*' (Prevaas, 2018). Typically, programmes integrate multiple interdependent projects to realize benefits unattainable through independent project management (Ferns, 1991). Within the program, employees work together on projects that are related to the overarching program, which is headed by a programme manager who oversees the goals and direction of the different sub-projects. The program manager makes sure that the sub-projects are in accordance with the overarching program goal and that programmes have enough resources. A key characteristic of programmes is they circumvent the normal hierarchy existent in organizations and instead facilitate cross-divisional work (Pellegrinelli, 1997).

Within the "Less Greenhouse Gases" program, the ILT brought together employees from different organizational units to co-create novel tools or ways of working. The researched project focused on developing an ML-based inspection tool (ML tool), with a project team of six core members. Throughout the project, a new end-user partially joined the group. The project team comprised members of either the Information, Networking and Programming Directorate or the Supervision and Investigation Service Directorate. Both directorates have different goals within the organization. The Supervision and Investigation Service conducts surveillance and oversight, for example by inspecting objects or companies. The Information, Networking and Programming directorate supports these activities by leveraging information and innovative ways of working.

Two project members enlisted themselves to be project managers within programmes. Before becoming project managers, they were also part of the Supervision and Investigation Service directorate. Throughout the ML tool's development, the project group met periodically. Between these meetings, project team members balanced their regular departmental responsibilities with project-specific tasks. The project managers were accountable to the program manager, with whom they also had various meetings.

The project group developed an ML tool that generates risks scores for potentially fraudulent companies by analysing their online marketplace advertisements. Designed for inspectors, the ML tool supports one of their responsibilities of conducting physical inspections, by identifying high-risk objects or companies. The ML tool was developed through the following process. First, a web scraping pipeline extracts online advertisements for fluorinated greenhouse gases. Then, stakeholders, primarily end-users, label these advertisements. The labelled data trains the ML model to distinguish fraudulent from legitimate advertisements. Then, when presented with new data, the tool predicts advertisement 'riskiness'.

An advertisement with a high predicted risk score indicates potential regulatory violations. For the inspectors, the risk scores serve as proxies for violations by the company that created the advertisement. The scores enable the inspectors to decide for follow-up research or physical company inspections.

4.3.2 Data collection

We collected data over a period of 12 months, from March 2022 to February 2023. During this period, the first author observed on multiple occasions and took part in some work at the ILT's innovation department, the IDlab. Part of the Information, Networking and Programming directorate, IDlab comprises mostly people with a data-science background dedicated to developing innovative, data-driven solutions that allow the ILT and end-users to make more informed decisions.

The first author was in close contact with the data scientists who were part of the ML tool project . They shared the ML tool and the issues that they encountered during the development and during their interactions with potential end-users. Although the innovation department was the first point of contact for the first author, access to various parts of the organization was possible.

The first author also took part in broader organizational activities. First, he participated in online ID-lab team meetings, in which employees shared various issues and future projects. Furthermore, he attended multiple physical meetings, like the yearly ILT festival, a day of organizational collaboration, team building, and knowledge exchange. Also, he attended a IDlab event connecting innovation-minded stakeholders across departments and directorates. Finally, the author participated in a meeting discussing the ML tools outcomes and the potential for application into new oversight domains. The interactions provided the opportunity for rich contextual observations of developer end-users interactions and ML

tool development. Additionally, they gave the opportunity to interact with both stakeholder types.

We triangulated the observations and presence at various on- and offline meetings with secondary data sources such as year reports, (data) strategy documents, frameworks for ML development, and a slide pack about the ML tool and an organization matrix. Triangulation in data collection increases the reliability and credibility of the results (Yin, 2013). The documents provided the necessary context information, and the strategy and organization surrounding the interactions between developers and end-users. Table 7 below gives an overview of the data sources and their use in the analysis.

Table 7: Data sources and use in analysis

Data sources	
Data type	*Use in analysis*
Primary data	
14 semi-structured interviews	Contributed to and enriched our understanding of the worlds of the stakeholders involved and the differences and issues they experienced while working with others.
7 hours of presence at online meetings	Contributed to and enriched our understanding of the background and ways of working of different stakeholders.
18 hours of presence at physical meetings	Provided broader insight into the workings of the organization, the background of the work of inspectors and data scientists, and the issues that emerge during collaboration.
Secondary data	
4 strategy documents	Provided insight into the goals of programme management and the overall organizational strategy.
5 year-reports	Provided insight into the development of the programme over time.
2 development framework descriptions	Provided insights into the way of working of developers and how they aimed to involve end-users.

Table 7: Data sources and use in analysis **(continued)**

Data sources	
Data type	*Use in analysis*
1 slide pack about the ML tool	Provided insight into the characteristics of the ML tool.
1 organization matrix	Provided insight into the overall organizational structure.

The first author conducted fourteen semi-structured interviews with fifteen participants. These participants were in the project group (seven), connected to that group, or connected to the program. Fourteen interviews were conducted, with privacy experts interviewed jointly. Table 8 presents an overview of interviewees.

With permission of the interviewees, voice or video recordings were made. The first author transcribed the recordings verbatim afterwards. During the interviews, he took detailed notes, which he re-ordered into a point-wise summary. The participants were selected based on theoretically driven within-case sampling. Hence, we purposefully searched for and contacted people who were potentially able to bring a novel perspective to the case based on their roles and positions within the organization. However, these people still needed to have an adjacent role in the Less Greenhouse Gasses program or the specific ML tool development sub-project. The interview actors included data scientists, inspectors, managers, project managers, and privacy experts. Speaking to persons involved on multiple different sides and with different responsibilities in the project increases the validity of the findings.

Table 8: Overview of interviewees

ID	Interviewee role
I1	Data scientist
I2	Data scientist
I3	Manager
I4	Inspector
I5	Inspector
I6	Privacy expert
I7	Inspector
I8	Manager
I9	Data scientist
I10	Data scientist
I11	Manager
I12	Manager
I13	Project manager
I14	Project manager
I15	Privacy expert

The semi-structured interviews had a degree of predetermined order, but also provided flexibility (Longhurst, 2010) and the possibility to go into topics that the interviewee indicated as important. The interview design allowed for exploration and follow-up on responses by the participants. The main themes, loose structure, and questions for the interviews were written down in an interview protocol.

This protocol included questions about the general role of the interviewee, their role in the project, their thoughts on the project, the choices they experienced working on the project, their future aspirations with the ML tool, and the general role of innovative ML tools in their work. Specific questions that were asked to every interviewee were: What is your role? What makes your work complex? However, due to the various and unique backgrounds of the stakeholders, the interviews allowed for exploration tailored to each stakeholders background and experiences.

The data scientists were asked about their development choices and ways for end-user participation. Inspectors provided insight into their experiences with data scientists, how much they felt involved in the development, and the operational implications of novel technologies. The project managers were asked about the interaction dynamics between data scientists and end-users, drawing from their unique position to facilitate and see the interaction. During managerial interviews we asked about strategic choices on various levels in the hierarchy, and explored how these choices could affect boundaries between developers and end-users. Lastly, the privacy experts contributed a valuable perspective, as they could be asked about their experiences with both inspectors and data scientists, and the boundaries that they experienced in working with both within and outside of the ML tool sub-project.

4.3.3 Data analysis

Our data analysis already started during the data collection. During the data collection, the first author regularly reflected on observations and linked these to related literature. Following the interviews, the coding process started. The first author conducted coding with ATLAS.ti, qualitative data analysis software. The first and second author frequently met up to reflect on the coding and to provide novel input to the codes. The initial phase involved reading the interview transcripts and field notes. While reading the interview transcripts the first author wrote down potential codes and highlighted key statements exemplary for the interviews. This helped to identify important themes. For example, our interest was sparked by the observation that many interviewees discussed the difficulties in collaborating with stakeholders who had divergent roles. This led us in the direction of understanding why these difficulties were present.

Afterwards, we performed open coding on the interview transcripts, which means that the coding followed the data. We performed this coding to identify emergent themes in the data (Williams & Moser,

2019). The data suggested that a reason for collaboration difficulties were the many differences between stakeholders throughout their work together. Informed by literature on knowledge sharing (Bechky, 2003) and managing knowledge across boundaries (Carlile, 2004) we then coded differences between stakeholders while co-creating and initially categorized these in various categories, such as 'language differences', 'expectation differences' and 'knowledge differences'. However, our coding also highlighted that many stakeholders mentioned ways that helped them collaborate more successfully with stakeholders with diverse backgrounds. We then carried out multiple rounds of axial coding, which relates to further refining, aligning, and categorizing the emergent themes (Williams & Moser, 2019). Furthermore, axial coding helps to capture relationships between multiple codes.

Especially in this phase, constant reiteration took place. We noticed that factors inhibited or promoted the successful blurring of boundaries. We realized that the data explained that boundaries persisted but also how boundaries might be blurred. At a later stage, we realized that both categories were linked to co-creation through programme management. Then, the mechanisms emerged through which programmes either allow boundaries to blur, persist, or emerge. Finally, the coding was checked for consistency and overlap. After coding, coded pieces of interview transcript that represent the codes were translated. These quotes are presented as statements from interviewees throughout the findings.

4.4. Findings

This section discusses the study's findings, categorized into three sections. Section 4.1 discusses how programme management allows boundary blurring. Section 4.2 discusses how programme management allows boundaries to persist. The third section discusses how programme management allows boundaries to emerge.

The findings are discussed as mechanisms that take place. While the mechanisms are described separately for analytical clarity, they are intertwined and form an intricate web that allows for boundaries to be blurred, to persist or to emerge. Figure 6 synthesizes the different boundary dynamics and mechanisms, and highlights how the different mechanisms impact different boundary types.

4.4.1 Programme management allows boundaries to blur

The findings indicate the existence of multiple mechanisms through which programme management allows boundaries to blur.

Selects the "right" end-users: Our findings indicate that programme management selects the "right" end-users. Programmes have the reputation of the place where innovations and change happen: *"And for innovation, that's what programmes are for" [I11].* This reputation is a reason some end-users are eager to join a programme. Because of its reputation, the programme 'selects' end-users that have a predisposition towards crossing boundaries. These end-users express to their management that they are eager to join the programme. Often such an end-user is someone who has a positive and realistic stance towards technology. They are intrinsically motivated to join the program because they are interested in innovation and change. These end-users are less invested in their domain-specific knowledge and already possess some level of knowledge about and interest in innovation. Hence, the semantic, syntactic and pragmatic boundaries that have to be crossed are relatively small compared to the boundaries between 'regular' end-users and developers. About such end-users it is mentioned that: "*The nicest thing is if you have positive people around it as well because otherwise, you don't get it done" [I7].* The interviewee conveyed the message that for projects that revolve around ML development blurring boundaries is heavily dependent upon the end-users that are involved from the start. This has been emphasized by a data scientist: "*You have to have people who really believe in this in advance" [I2].* These people are end-users

who are intrinsically motivated to contribute to innovation and are open to crossing boundaries with developers of such technologies. Within the program, these envisioned end-users are provided space for their willingness to explore and innovate.

Additionally, the right end-users can blur boundaries outside of the programme. These end-users can function as a mediator:" I *also ended up being a kind of mediator at some of the consultations, making small presentations about the tool'" [I5].* The inspector mentions consultations with other colleagues and future users of the ML system who were not directly involved in the programme. The project team organised consultations with end-users outside of the programme to present the results of the project. The end-users that had been involved mediated between the developers and the more sceptical and less involved potential future end-users. They spanned the boundary between two occupational communities. The right inspector can even function as a kind of 'owner' of the ML system within the group of inspectors, whenever the ML system has been taken up by the team. The inspector might function as *"A coordinator of the outputs from this tool. The inspector ensures that the work for this is done" [I14].* The existence of the programme helps to get the involvement of an end-user with a techno-optimist stance towards ML technology and the abilities and motivations to function as a mediator or coordinator. The right end-user can function as a bridge between both worlds. Without the program, there are fewer opportunities to build bridges.

Stimulates open-mindedness. Our findings indicate that programme management stimulates open-mindedness for stakeholders who are part of a project. Programme management and the projects that occur within the context of programmes are characterized by more freedom and experimentation: "*Incidentally, we have the freedom in a programme to experiment" [I11].* Whenever developers and end-users are part of the program, they can "*Stand away from it, stand above it "[I11].* This

refers to the developers' and end-users' activities and way of looking at their normal day-to-day work. Their day-to-day work, which happens in the context of their respective teams and departments '*in the line*' is characterized by a distinct perspective. This perspective entails reaching goals that are part of the team's yearly plans, taking common actions, and a way of looking that accompanies those actions and goals. The programme provides the opportunity to take a step back from the line and the way of looking that is stimulated under that banner. The programme allows to temporarily detach from a community of practice and become part of a project and novel community that is not yet set in stone. In some way, both occupational communities start from a somewhat blank slate in terms of practices and the used language. The project gives the opportunity "*to interpret from a broader perspective" [I12]*. This also relates to the ambiguousness and complexity of the project. Whereas goals outside of the program are often clear cut and set in stone, goals within the program can be subject to change and the underlying problem to solve as well. Hence, the programme allows the developers and end-users to temporarily free themselves from the mindset that is stimulated in their day-to-day work. Instead, programme management provides the opportunity for developers and end-users to be part of multi-disciplinary project groups, in which these stakeholders can co-create new perspectives in a more open and ambiguous environment.

Respects the cyclical nature of ML development. Our findings indicate that programme management, due to its characteristics, respects the cyclical nature of ML development. It respects a way of working that many developers already practice. The findings are supported by multiple interviewees (I7, I10, I11, I12). The cyclical nature of the program was highlighted by a manager responsible for the program: "*It is important to me to have clear progress reports in the meantime.*" [I11]. These progress reports or progress meetings involve the programme manager and the managers who are responsible for the project. They provide the opportunity to give the project a new direction. In essence, there are

multiple moments where decisions regarding the direction of the project can be made. Developers acknowledge that the cyclical nature of ML development is respected. Also, because the development of ML happens within this programmatic context it becomes easier to engage people throughout the different cycles: "*So not only the result but make sure you keep people hooked in the process, maybe get them more motivated*" *[I12].* Programme management closely fits with existing ML development approaches. This is something that developers are used to. However, outside the program context it is hard to respect that process. But, within the program it becomes easier to organize the ML development in a cyclical way, thereby blurring semantic, syntactic and pragmatic boundaries. That programme management respects the cyclical nature also was mentioned by end-users: "*This is something you should do in cycles. Discuss, look at what you have built and be critical of whether you need to go further or not" [I7].* Hence, also the involved end-users highlight that co-creation an ML tool works best in a cyclical way.

Encourages shared performance and goal setting. Our findings indicate that programme management encourages shared performance and goal setting. The program helps to overcome pragmatic boundaries because it encourages to work towards common interest. The mechanism partially comes into play because being in a programme takes time. This has been emphasized by end-users in the project. An inspector mentioned: *"We have already put our scarce time into this. What if we have zero results*" *[I7].* The end-users and developers are often part of a project within a programme because they want to be. However, this does not mean that their other obligations are put on hold. Thus, there is a strong wish and incentive to use the time dedicated to the program efficiently. This is also something that one of the project managers noted about the involved developers and end-users: "*Basically they all indicated: I do want to help you, but then I have to be sure that my work leads to results*" *[I13].* Hence, there is a strong wish present in both for actual performance, or said differently, tangible results. Such a tangible result -. i.e. a 'product' - can

serve as a boundary object, bringing together developers and end-users around it. The findings show that shared performance can be encouraged through goal setting. One of the involved end-users mentioned: "*Of course, we said what we hoped to get out of it, but I think in cautious language" [I7].* The involved developers, end-users and project managers responsible for the project together discussed what the results of the project should be. The group negotiated towards a shared, tangible and clear goal, that they then tried to reach in the brief time available to them. With clarity about the potential result, their time investment became more tangible and boundaries became blurred.

Enables shared ownership: Our findings indicate that programme management enables shared ownership between developers and end-users. Proximity is seen as a condition for shared ownership. A manager stated: "*In my opinion, a programme's strength lies in its ability to bring together all those different roles in this way" [I11].* What the interviewee refers to is the fact that in the context of programmes different stakeholders in the development of ML technologies can foster close relationships. Such close relationships are less enabled when working outside of the programmes. This is further emphasized by a project manager who mentions: "*And that's sort of what these programmes are for. To pull people from different line departments together" [I14].* Within a programme, occupations work together on temporary projects. These temporary projects enable shared ownership as developers and end-users work closely together on a goal that they have decided together. In a way, while working in a program, these occupations temporarily can detach from their obligations 'in the line' and come together in close collaboration. Pragmatic boundaries are partially overcome, because developers and end-users are enabled to become shared owners of a project. Both groups feel a responsibility to move the project forward.

The findings also indicate how such shared ownership plays out in practice. The ownership happens because of a proximate collaborative

face-to-face relationship between developers and end-users. Both groups of stakeholders meet each other physically and work together on the tool. Both sides provide input and make decisions about steps that need to be undertaken or data or considerations that must be incorporated. One of the interviewees mentioned: "*I personally find the added value of this project that we brought technicians and inspectors together from day one" [I14].* The interviewee refers to the developers (technicians) and end-users (inspectors) and indicates that bringing both groups together from the beginning has been instrumental. Likewise, another project manager added: "*It has been a co-production"[I13].* The project manager indicates that co-production (i.e. shared ownership) was one of the reasons for the success of the project. The essence of the collaborative and proximate relationship is a non-hierarchical collaboration, in which both occupational communities bring their expertise to the table. This has been acknowledged by inspectors: "*I think we really developed it together" [I7].*

4.4.2 Programme management allows boundaries to persist

The findings indicate the existence of multiple mechanisms through which programme management allows boundaries to persist. Hence, here we discuss findings that deal with what program management cannot do and why that is the case.

Allows for persistent language differences. The findings indicate that programme management allows for persistent language differences. These language differences are related to the lexicon that developers and end-users have , and the meanings they attach to certain words or terms. Hence, both semantic and syntactic boundaries persisted. This is important , because the co-creation of a useful ML tool demanded the input of end-users. Their input is needed to label the data. If language differences persist, the data quality can suffer, thereby impacting the quality of the ML tool. While the stakeholders met on various occasions and had the opportunity to develop a shared understanding,

the temporary nature of the project work and the diverse backgrounds made it hard to create a shared language. As was mentioned: "*So you have to make those translations all the time " [I10].* This data scientist explained that they constantly had to find the right words that would make it understandable for the end-users, despite both occupational communities working together in a context of co-creation. The data scientists had to make sure that they empathized with the viewpoint and language of the end-users. That different meanings persisted can also be illustrated by a statement made by an end-user: *'Because everyone calls it a tool, except the people who made it and I don't know what word they use for it'"[I5]*. The statement illustrates that despite talking about the same 'thing' (i.e. the ML tool) both occupational communities use different words. In essence, the co-creation provides boundaries around language to partially blur, but some will always remain. The program can only do so much. It is only temporarily, and there are no formalized ways to developed a shared language. Hence, the freedom also has a downside. Furthermore, in times of uncertainty, occupational communities might fall back on what they already know, thereby keeping differences and boundaries intact.

Allows for persistent knowledge differences. The findings indicate that programme management allows for persistent knowledge differences. It was mentioned that *"Data science is like magic for a lot of inspectors." [I2].* The data scientist indicated that the 'average inspector' has a limited level of knowledge regarding what ML is and how it could help end-users. These knowledge differences seem to persist because the project was focused on tangible results (also see *Encourages shared performance and goal setting).* This focus on tangible results also strongly ties into the time pressure that especially potential end-users experience. A focus on results and having tangible products along the way might disregard the basis that various stakeholders reason from. Co-creation through programme management cannot change that both groups of stakeholders have a strongly different knowledge basis. Secondly, as noted by one of

the interviewees: "*They have no idea what an inspection looks like in practice*" [I4]. This end-user referred to data scientists and the knowledge they have about the activities of inspectors. Programme management brought developers and end-users together. However, they were not present in each other's day-to-day working practices and processes. Instead, they met each other in a context of experimentation and open-mindedness. This helped them to develop a shared basis. However, at the same time, due to the temporary and ad-hoc character of their work together, they were unable to create common knowledge about each other's activities. Likewise, a data scientist mentioned: "*We do not have the substantive knowledge about understanding what is actually happening" [I2]*. The data scientist indicated that it was impossible to have the inspectors' deep base of knowledge based on their experience and obtained tacit knowledge. They are unable to fully understand the reality the end-user sees. Programme management temporarily provides the opportunity for co-creation, but this organizational intervention cannot overcome knowledge boundaries based on years' worth of implicit knowledge obtained through inspections and embedded in the practices of end-users. Also, inspectors and data scientists tend to label each other's knowledge to make it more tangible. An inspector mentioned: "*We always call it knowledge of content and they are the technical knowledge*" [I7]. The statement suggests that data scientists may be viewed solely as technicians, while end users may not possess the same level of technological expertise or interest. The statement emphasizes the different perspectives of developers and end-users. Also, it shows how framing each other's roles might leave boundaries intact, or even reinforce them.

Enlarges time and resource differences. Our findings indicate that programme management enlarges time and resource differences. The context of programmes is one of exploration and space for innovation. Being innovative and working on innovative technologies is one of the main tasks that data scientists are hired for. When these developers are

part of projects in the context of programmes, they are given the time and space (by their line managers) to work on the program. It is congruent with their main task in the organization. However, end-users are dealing with a different reality. All interviewed inspectors are experiencing a lack of resources and time pressure. As an inspector noted: *"It's just continuously busy at the moment" [I7]* and another inspector: "*We just don't have the capacity to do something experimental right now" [I5].* Also, both project managers, who were responsible for the ML tool project, indicated that they observed a strong lack of time and resources from the end-users. They mentioned: "*Ultimately, it is about resources and perhaps a lack of time or a lack of means" [I14].* The end-users are hired to inspect companies, take the necessary actions based on their findings, and have an overview of their oversight area. However, working on innovative tools in the context of programmes interferes with their daily work and goals. Often, these end-users are assessed based on different measures rather than being innovative. Often, they like to dedicate time towards activities that will help them reach the goals on which they are assessed. However, for developers, programmes are an opportunity to dedicate even more time to innovation and exploration. This creates and enlarges the discrepancy between developers and end-users.

4.4.3 Programme management allows boundaries to emerge

The findings indicate the existence of multiple mechanisms through which programme management allows boundaries to emerge.

Facilitates expectation differences. Our findings indicate that programme management facilitates expectation differences. This finding does not only relate to involved developers and end-users but also to end-users not directly involved in the programme. Programmes bring together stakeholders from different departments. However, program management does not necessarily solve a discrepancy in expectations. Even more so, it provides the opportunity for involved stakeholders to project their own beliefs and perspectives on what program management can

and cannot provide. Some end-users have grand expectations of what programme management and working on innovative tools based on ML will provide them. A developer mentioned: "*They expected there to be golden mountains, but they are not there"[I2].* The statement illustrates how some end-users had completely different expectations than what the co-creation resulted in. A statement from an end-user adds to this*: "I expected that we would suddenly come across companies that we did not know at all…And that is not what happened*" [I4]. This illustrates the disappointment this end-user experienced about the results of the programme and the ML tool. Despite the existence of the programme, the expectations were not aligned. This might be partially caused by what program management is not. Program management is a temporary way of working together, often cross-divisional. But it also somewhat vague, providing space for differences in expectations. This misalignment in expectations can even be observed between end-users:" They *(other inspectors) had expected that the tool would solve it. And I thought, this makes our job easier" [I5].* This statement illustrates how programmes, because of the differences in involvement of different end-users can contribute to novel boundaries. Expectation differences can be significant for knowledge boundaries if we assume that overcoming such boundaries requires much effort. Expectation differences among employees create novel boundaries between the willing and unwilling to make this effort.

Facilitates differences between types of end-users. Our findings indicate that programme management facilitates differences between types of end-users. An end-user indicated: *"And there's just really a difficult dividing line and I think you only have that on the user side" [I7].* Being part of a program means that end-users are temporarily not only working on the targets and tasks from their departments but also working on program-related tasks and goals. Hence, they become part of a novel community and share and build knowledge inside that community with different stakeholders. This also means that the potential end-users that are outside of the program are less involved in the development of the

ML tool. In the end, not only end-users involved in the project team but also those outside the program are expected to use ML tools. When end-users in the project become too close with their project team and developers there is a risk that they lose touch with end-users outside of the program. The differences between these groups grow, which makes it harder to eventually transfer a potentially successful ML tool into the organization. Thus, for end-users there is a risk of being too close with their project team: "*I think as a project team we were really close" [I7]*. This can hinder the transfer of an ML As such, co-creation is a double-edged sword tool into the broader group of end-users as they feel even less connected to what their colleagues are doing. This has consequences for the organization. The top management of the organization hopes to gain valuable perspectives and novel ways of working from working in programmes. The idea is that these perspectives and ways of working are continued in 'the line': "*You just want continuity of things that work and are good" [I12]*. However, when differences between types of end-users are facilitated, or even enlarged, they might hinder the eventual use of potentially helpful ML tools. Key to this finding is the quality of programme management. The emerging language on performance and results, as much as the emerging ownership serve as bridges over the boundaries between developers and end-users. At the same time, they lead to differences between end-users who take part in programme management and those that did not participate. The latter did not invest in developing the shared language and are not owners of the results.

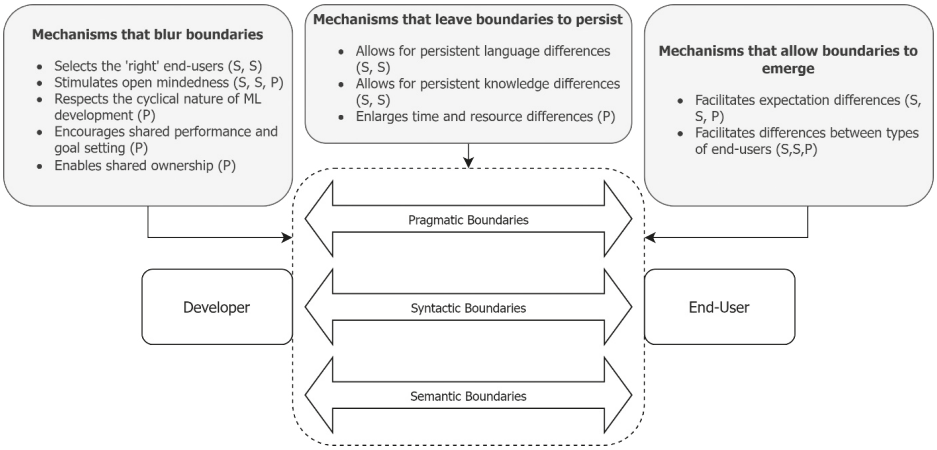


Figure 6: Overview of boundary dynamics and mechanisms

4.5. Discussion

Our study seeks to address the question of how co-creation influences boundaries between developers and end-users in ML development and how boundaries behave as a result. Table 9 highlights the three dynamics that we observed – blurring, persisting, and emerging – and the effects that occur at the different types of boundaries. In presenting these findings as such, we took inspiration from Barrett et al. (2012) in their presentation of boundary effects in pharmacy work.

Below, we will elaborate on the main findings in more detail and highlight the contributions to different areas of the literature. First, we discuss how co-creation can have unexpected opposite effects by causing novel boundaries to emerge, referring to this phenomenon as the 'waterbed effect'. Second, we elaborate on how boundaries might persist, while giving the appearance of being blurred. This leads us to highlight the layering of boundaries. Third, we theorize that pragmatic boundaries in co-creating ML are partly the result of differences in goals and incentives coming from the original departments of developers and end-users. These differences lead to a skewed distribution of transaction costs.

Table 9: Description of boundary dynamics and boundary effects

Boundary dynamics	Boundary effects
Blurring	· Blurring of the boundary between occupational groups · Increasing the knowledge base of occupational groups and reducing the distance between knowledge bases
Persisting	· The lack of impact on deep-rooted boundaries leaves boundaries between occupational groups to persist · Persistence of syntactic and semantic boundaries instead of pragmatic boundaries (layering)
Emerging	· Measures to blur boundaries have unexpected opposite effects by causing novel boundaries to emerge (waterbed effect) · Boundaries within occupational groups trough divergent expectations and involvement

4.5.1 The waterbed effect of co-creating AI

The introduction of programme management brought developers and end-users closer together. This environment facilitated shared ownership, shared goal-setting, and open mindedness. In providing this environment, the developers and end-users were facilitated in overcoming and blurring pragmatic boundaries. The knowledge of both was less 'at stake', because developers and end-users were expected to work towards a common goal. As both were temporarily detached from their normal practices space was offered to form what has been called 'a new joint field' (Levina & Vaast, 2005). As such it seems that co-creation was successful in blurring boundaries between developers and end-users. However, the intervention was only partially 'successful', as new boundaries emerged as well.

Hence, a waterbed effect occurred, that is depicted in figure 7. In the context of ML development and boundaries it refers to the undesirable emergence of novel boundaries elsewhere, while trying to blur boundaries in another place. Boundaries were blurred between developers and end-users. At the same time, novel boundaries between different types of end-users emerged. While the program runs, developers and end-users make joint efforts towards a shared goal and create new language within

the process. However, when developers and end-users then return to their teams, especially the end-users who were involved experience emergent boundaries. The programme's goals that developers worked on and the language that was formed are far removed from the end-users original practices. Hence, novel boundaries between end-users have emerged.

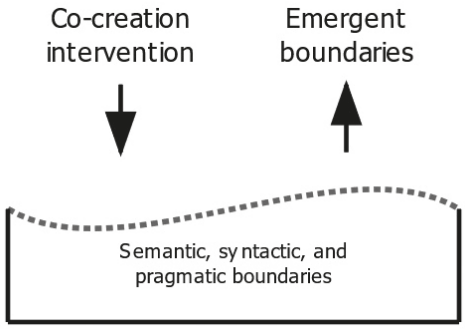


Figure 7: Illustration of waterbed effect

That novel boundaries can emerge as a consequence of organizational interventions aimed at blurring boundaries has also been highlighted by Waardenburg et al. (2022). Our findings substantiate this claim within the specific context of ML development in public organizations in two ways. First, we show that an intervention with the explicit aim to connect occupational groups may lead to emerging boundaries somewhere else. Second, we show that unexpected effects may even occur when developers and end-users are in direct contact with each other, instead of only through a liaison position, as in the work by Waardenburg et al. (2022). These findings challenge the manageability of boundaries. While interventions to connect multiple groups might appear effective on the level of the focal community, a broader systemic perspective reveals the emergence of novel boundaries. More broadly, our finding adds to the literature that calls for the need to blur boundaries, and highlights how interventions aimed to do just that play out in practice (Mayer et al., 2023; Waardenburg & Huysman, 2022).

The findings call for more research on the dynamics of knowledge boundaries. An exploration of causalities between boundary blurring and emerging could be promising. Theorizing may suggest that 'waterbed effects' are likely to occur when key resources – such as time - remain constant. In our case study, a part of the end user's community was able to engage in a joint learning process with developers, accidentally creating a new boundary with the remaining end-users. Empirical research testing these causalities may provide valuable insights into the distribution of novel knowledge. Such an investigation could illustrate how novel knowledge – in this instance, ML expertise disperses across an organization and at which pace.

4.5.2 The layering of knowledge boundaries

A second area of contribution concerns the understanding of boundaries as they appear in ML development. As noted, through co-creation developers and end-users were brought closer together and worked towards the same goal. On the surface it seemed as if semantic, syntactic, and pragmatic boundaries were overcome. However, as the process carried on it became clear that semantic and syntactic boundaries were more entrenched than expected. The developers had to put effort into translating machine learning concepts in words that the end-users would understand. However, even if this effort seemed successful words proved to have different meanings among the communities. It seems as if semantic and syntactic boundaries were more fine-grained than expected and blurred slowly, while pragmatic boundaries seemed to be overcome quite quickly. As such, co-creation can give the appearance of syntactic and semantic boundaries being blurred, but they may implicitly remain.

These observations imply that different types of boundaries may change in a different pace. As such, we suggest that knowledge boundaries can be seen as layered phenomena. Boundaries are not stand-alone and homogeneous entities, but a combination of multiple boundaries that

interact. Some boundaries may persist while others blur. Some boundaries blur, while others persist implicitly.

The interplay between boundaries and the different 'layers' that exist have also been described by Carlile (2004). It is mentioned that the more complex boundaries become - with syntactic boundaries being the least complex and pragmatic the most – that '*the process or capacity at a more complex boundary still requires the capacities of those below it* ' (Carlile, 2004, p.560). This suggests that for blurring pragmatic boundaries, the blurring of syntactic and semantic boundaries is a prerequisite. However, we have not found blurred syntactic and semantic boundaries to be a prerequisite for blurred pragmatic boundaries, as semantic and syntactic boundaries actually persisted, while pragmatic boundaries appeared to be blurred. This could be explained by an implicitness of syntactic and semantic boundaries, as they are grounded in both wording and meanings. The wordings seem measurable, the meanings to a lesser extent. This difference may be significant in the case of ML, a technology that not only introduces a new vocabulary, but also a novel way of thinking about the work processes, in our case risk-based oversight.

The observation that pragmatic boundaries seemed to blur quicker than semantic and syntactic boundaries may be subject to case selection bias. To blur pragmatic boundaries was more or less the programs' aim. Developers and end-users were intentionally united through co-creation goals. However, the case highlights the tenacity of semantic and syntactic boundaries, and the possibility that different boundary types may blur independently from each other. More empirical research could explore this relationship, for instance studying interventions aimed at blurring syntactic, semantic and pragmatic boundaries separately.

4.5.3 Work processes as source of pragmatic boundaries in ML development

The last area of contribution is the understanding of the source of pragmatic boundaries in ML development. Carlile (2004) notes that a pragmatic boundary arises when knowledge cannot be shared because of different interests among actors, thereby introducing a political dimension to knowledge sharing. Carlile (2004, p.559) states that costs "*are not just the costs of learning about what is new, but also the costs of transforming "current" knowledge being used (i.e., common and domain-specific knowledge)*". These transaction costs negatively affect the willingness to share knowledge. Carlile (2004) also discusses that we should not assume the actors involved at a boundary occupy politically equal positions in representing their knowledge to each other.

Our research reveals that while shared ownership and common goal-setting initially seemed to blur pragmatic boundaries, looking beyond the surface exposes transaction costs as pointed out by Carlile (2004). We also observed an uneven distribution of transaction costs in ML development. End-users encountered significant obstacles in the form of impenetrable ML development language and unfamiliar working processes. The cyclical nature of programme management aligns more closely with ML development processes and the workflow of developers then of end-users.

This misalignment leads to asymmetry. End-users must adapt to a novel reality that is fundamentally different from what they are used to. Developers, on the other hand, operate within familiar territory aligned with their innovation-focused objectives. The political dimension of pragmatic boundaries in ML development thus emerges from these inherent work process disparities. This is also a key contribution to the literature that highlights the need for interaction (Waardenburg & Huysman, 2022) as it highlights how unequal positions may lead to interaction issues. ML development presents additional challenges

beyond standard knowledge transformation. Its data-driven, algorithmic nature represents not only a language boundary, but may also demand fundamental epistemological shift from end-users.

These points prompt critical future research avenues. First, we suggest that studies question the legacies that determine the (un)equality between actors that want to overcome pragmatic boundaries. How is the distribution of transaction costs among these actors determined by their work processes and other legacies? Does the distribution bias apply to all innovation processes, and would the impenetrability of ML to end-users as well as a change of epistemology lead to an amplification of this distribution bias? Empirical work focusing on these questions is promising, since it may help to predict power distribution among ML developers and end-users.

4.5.4. Limitations and further research perspectives

While our findings offer insight into boundary dynamics due to co-creation, several limitations need mentioning. First, our single case study design may limit external validity of the findings. Second, selection bias may influence our findings, as the studied organization seemed mature in their ML implementation. Hence, this is likely to have contributed to the openness of this organization to research participation. Third, our focus on interactions between developers and end-users excludes other crucial occupational communities needed in ML development, such as legal experts and ICT architects. This is a limitation because ML is a technology of many hands (van Krimpen et al., 2023; Zweig et al., 2018).

Future research opportunities are manyfold, as discussed above. First, we suggest studies into causalities between boundaries blurring and emerging. Also, we suggest studies that focus on the (inter)dependencies between different types of boundaries and how interventions interact with different boundary types. Next, we suggest research on the distribution

bias of transaction costs. These future studies could reveal new dynamics around waterbed effects, layered boundaries and transaction costs.

4.6. Conclusion

This study aims to answer the question: *How does co-creation influence boundaries between developers and end-users in ML development and how do boundaries behave as a result?* The study suggests that co-creation has mixed and unpredictable effects on boundaries between ML developers and end-users. While co-creation has various mechanisms to blur boundaries, like selecting the right end-user and enabling shared ownership, our study demonstrates the persistence of boundaries. Indeed, co-creation interventions like programme management may generate a 'waterbed effect', where blurring boundaries in one place, triggers the emergence of boundaries in another place.

Our investigation of boundaries and boundary dynamics showed an interplay of characteristics of boundaries and a layering of boundaries. We found differences in pace between the blurring of distinct types of boundaries. Some mechanisms that potentially blur boundaries emerge quickly, while others tend to persist longer. These mechanisms, like shared ownership, may allow for conditions to cross more persistent boundaries in the long run. Counterintuitively, in ML development co-creation, pragmatic boundaries seem overcome faster than semantic or syntactic boundaries. The findings call for a layered conception of knowledge boundaries, wherein old and new practices co-exist and interact. Finally, our analysis reveals that blurring knowledge boundaries through co-creation is characterized by the presence of transaction costs, stemming from work processes in program management that characterize ML development processes.

This study emphasizes the message by Beckhy (2003) once more. She illustrated that '*interactions between members of different communities, while sometimes painful, can lead to enriched understanding, particularly when a tangible object is offered to ground the interaction*.' Also, in the case of the co-creation of ML, there seems much to gain by seeking ways to increase the interactions between developers and end-users in an open, collaborative, and equal manner. However, the 'waterbed effect' tells us that there is no world without boundaries. Given that knowledge boundaries may impede organizational learning, the motivation for co-creation interventions remains clear. However, for these interventions to be effective, we call for more understanding about how co-creation interventions impact boundaries and how boundaries behave and interact.

Acknowledgements

The authors thank handling editor Michael Barrett and the anonymous reviewers for their valuable feedback and guidance during the revision process. Furthermore, they thank the ILT, more specifically Stephanie Wassenburg and Corline Koolhaas, for their supporting role in data collection.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used DeepL and Claude in order to perform proofreading and editing on the written manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

'Mama always had a way of explaining things.'
– Forrest Gump in movie: Forrest Gump (1994)

5

Explainable AI as a Learning Process: The Impact of an Interactive and Iterative Development Process on XAI for Organizational Stakeholders

This chapter is based on a revised manuscript submitted to Government Information Quarterly Special Issue: AI in Government: Design, Implementation and Oversight.

Abstract

This study investigates how interactive and iterative AI development affects explainability (XAI) through action research in a public sector organization. While XAI research increasingly recognizes the importance of social elements and stakeholder interaction, the impact of interactive and iterative development processes on explainability remains unexplored. Our findings demonstrate that such development increases explainability because they instigate and facilitate organizational learning and highlight the importance of differentiating between internal stakeholders. This process significantly impacts developer roles, stakeholder involvement, and AI development practices.

Keywords
Explainable artificial intelligence; explainability; organizational stakeholders; interactivity; development process; learning processes

5.1. Introduction

Scholars have mentioned that face-to-face social interaction and informal relationships can build up the trust needed for the successful introduction of complex AI projects (Campion et al., 2022). The logic is that working in close physical proximity helps people learn from one another, and stakeholders from different backgrounds can establish trusted relationships, creating an environment where trust in the AI model can develop. Such work based on interactive and iterative AI development is closely related to the streams of literature that see AI as a configuration of involved people (Bailey et al., 2022; Schneider et al., 2023) who interact iteratively with each other within processes, and are embedded within organizations (Janssen & Kuk, 2016; Lorenz et al., 2022; Meijer et al., 2021). These contributions highlight the complexity of AI development and show that the many internal stakeholders must collaborate closely for successful AI development (Waardenburg & Huysman, 2022). According to this stream of literature, organizational stakeholders' trust in AI may be the consequence of AI applications being developed in a collaborative development environment, where diverse stakeholders work together.

Another way to create trust in AI among all stakeholders is through explainable AI (XAI). AI explainability is seen as a solution to 'black-box AI' i.e. the lack of transparency in certain types of AI (Belle & Papantonis, 2021) and can help to create trust in AI (de Bruijn et al., 2021; Meske et al., 2022). Nevertheless, it is hard to achieve explainability in practice, a point emphasized in the literature covering the different perspectives on explainability. A number of authors have discussed technical approaches that focus on explaining AI models (e.g. Barredo Arrieta et al., 2020; Belle & Papantonis, 2021; Bhatt et al., 2020). Recently, others have included a social perspective and highlight the need to differentiate between the viewpoint of end users, developers, managers and other stakeholders in XAI, who have different demands for explainability (Longo et al.,

2024; Preece et al., 2018; Tomsett et al., 2018). However, other scholars have highlighted that awareness about different stakeholders and their explainability needs is necessary, but they also indicate that the involvement of stakeholders within AI development can be a condition for explainability (de Bruijn et al., 2021). Interacting with stakeholders throughout the AI development and understanding their needs might lead to additional explainability (Bhatt et al., 2020)

Despite the growing understanding of the importance of the development processes behind AI, and the presumed need for interaction and iteration, there is still little empirical work on the impact of such a development process, and even more so within public sector organizations. Our investigation into the literature indicates that an interactive and iterative development approach could help increase the explainability of AI models (Bhatt et al., 2020; de Bruijn et al., 2021; Longo et al., 2024). The idea that an interactive and iterative AI development process can potentially contribute to achieving explainable AI has yet to be studied empirically within the public sector. Given that the potential benefits and challenges with this approach have only been highlighted theoretically, in our study, we explore this issue by answering the following research question: *What are the benefits and challenges of an interactive and iterative development approach for explainability of AI models and how do benefits and challenges occur?* We aim to give an overview of this interactive and iterative development approach and explain how it relates to explainability.

To address this question, we applied action research to an Italian public organization which was undertaking an elaborate AI project. Our study presents an overview of benefits and challenges of interactive and iterative AI development to contribute to explainability of AI models. The project was exemplary, being a feasibility study within the early development stages of an AI application designed to augment a complex decision-making process.

The remainder of this article is organized as follows. Section 2 presents a review of related work. Section 3 provides an explanation of our methodology, and the findings are discussed in Section 4. Section 5 contains a discussion of the results and a comparison between related work and our findings. The conclusions are lastly set out in Section 6.

5.2. Related work

5.2.1. The emergence of stakeholder-centric XAI

Much of the traditional literature on explainable AI comes from computer science and similar fields. The earliest work on XAI focused on expert systems (de Bruijn et al., 2021), but more recent XAI literature has focused on explaining AI models that are inherently opaque, mostly referred to as machine learning models, which learn from data (Domingos, 2012; Du et al., 2020).

A key term is 'models' as this has mostly been the focus of XAI. Some scholars, like Tomsett et al. (2018) have taken a broader view, alluding to 'the system' as the element that should be explainable. They refer to a machine learning system as '*a system that includes one or more machine learning models, the data used to train the model(s), any interface used to interact with the model(s), and any relevant documentation*' (p. 9). However, they also encapsulate the term 'model' in their definition. This further highlights that explaining models was and is a key focus of the field of explainable AI.

Other scholars (e.g. Barredo Arrieta et al., 2020; Belle & Papantonis, 2021) discuss the detailed ways of how to make models explainable. Barredo Arrieta et al. (2020) prepared an overview of the subject, mentioning different types of post-hoc explainability techniques. Often referred to as post-modelling explainability, these techniques "*aim at communicating understandable information about how an already developed model*

produces its predictions for any given input" (p. 92). The authors mention explanation by simplification, feature relevance explanation, local explanation and visual explanation. This highlights that there is not a single way to explain 'a model'. Belle and Papantonis (2021) discuss the mentioned techniques in greater detail, describing, for instance, how explanation by simplification works. However, these techniques that are focused on 'the model' and its inner workings are not the only XAI techniques in existence.

Closely connected work in XAI has also focused on specific instances of input and connected output, rather than the entire AI model. A specific instance of input and connected output is for example the decision for a particular individual receiving a 'high' risk score. The distinction between explanations relating to the model versus explanations of specific input-output relationships (specific instances) is often discussed in the literature in terms of the difference between global and local explanations (Bhatt et al., 2020). A drawback of merely explaining local instances is that it only gives a limited picture of the system, meaning that the explanation underpinning a single instance may not be representative of the model decision-making process as a whole (Langer et al., 2021).

The literature discussed so far highlights that there is a wide variety of approaches available to explain models or specific input-output relations. However, these explainability techniques also have drawbacks. For example, besides the developers that developed the model, others may not have the appropriate knowledge to be able to understand the explanations (de Bruijn et al., 2021). Hence, the focus of XAI on explaining models or specific instances might not be sufficient for various stakeholders.

Consequently, in recent years the field has evolved from a strong focus on models towards a stronger emphasis on the various stakeholders surrounding AI models. Many authors mention that different stakeholders might require different explanations (e.g. Brasse et al., 2023; Langer

et al., 2021; Longo et al., 2024; Meske et al., 2022; Preece et al., 2018; Tomsett et al., 2018). Different scholars categorize the stakeholders in different ways. For example, some mention six 'roles' that stakeholders can have surrounding AI systems (Tomsett et al., 2018), while others mention five stakeholder groups and provide a visual tool suggests that AI managers are the central stakeholder group (Meske et al., 2022). Although categorizations differ, their existence highlights that scholars have thought about the various stakeholders surrounding AI models. Tomsett et al. (2018) also discuss why certain stakeholders may want to have a different explanation. For example, end-users are interested in knowing whether the AI's decisions are fair and if they can be challenged and overridden, whereas developers want to see if the models perform correctly against performance metrics. These scholars bring the attention to the fact that stakeholders have different explainability needs.

However, multiple scholars bring attention to the fact that even categorizations of stakeholders do not capture the complexity of various explainability demands (Langer et al., 2021; Longo et al., 2024; Meske et al., 2022). Langer et al. (2024) make the point that most stakeholders discussed are merely prototypical, but in reality there is no 'prototypical developer'. A novice developer will have different explainability demands than one with more experience. Similarly every stakeholders has a different background and brings other preferences and abilities that influence the explanation they require (Longo et al., 2024).

5.2.2. The implications of interactive development for XAI

Given the point made by several scholars that stakeholder categorizations for XAI have drawbacks and do not capture the complexity of the explainability demands (Langer et al., 2021; Longo et al., 2024; Meske et al., 2022), the focus within the field XAI has started to shift even further. Instead of explaining models, specific input-output relationships, and taking into account various stakeholders, the development process of AI

models can be a source of explainability. The development process might contribute to explainability in two ways.

Firstly, de Bruijn et al. (2021) observed that it can sometimes be hard to explain models. One reason they mention is that circumstances may change, for example a model may be re-trained with new data. These scholars noted that, instead of explaining the model, it may be necessary to explain the processes behind the model. Hence, they call for 'explainable processes', meaning that the processes behind the algorithm should also be explainable. These processes consist of the design processes and the debate about important decisions within that process. This perspective shifts the focus from the need to explain a model or system to the dynamic process behind that system. This more 'qualitative approach' potentially contributes to a richer explanation than can be achieved through technical approaches, but equally less 'measurable'.

A second way in which AI development processes might contribute to XAI is through the form and structure of that process. Hints about how the development process can contribute were given already in more technically oriented XAI work. For example, Bhatt et al. (2020) offered recommendations on how to achieve explainability for stakeholders. They recommend developers to engage with stakeholders and asking them about the purpose of the explanation. For example, does the stakeholder need a once-off explanation or must the explanations be dynamic? This work highlights that interaction with stakeholders could be vital to ensure explainability. Similar comments have also been made by Miller (2019) through framing explanations as social interactions. Similarly, de Bruijn et al. (2021), mention the potential for 'negotiated algorithms', which, as the name implies, are negotiated, or developed between several stakeholders through an interactive and iterative process. Through developing the algorithm together, the involved stakeholders gain understanding about the models and processes behind these models. Such work acknowledges

that explainability is often not necessarily seen as a property of a particular model, but as an ongoing process (Hong et al., 2020).

Despite the progressive understanding about the importance of the development processes behind AI, and the need for interaction and iteration, still not much empirical work is done about the effects of such a development process. It is thus not surprising that scholars still mention that '*designing and tailoring appropriate explanations for each stakeholder type, both in terms of content and format and presentation, is an ongoing challenge*' (Longo et al., 2024. p13). This paper aims to focus on the format of explainability, in this case an interactive and iterative development process. These principles, interaction and iteration, can be seen as guiding the development. In that sense, they also answer to the call by Meske et al. (2022), asking for more research into the principles on how to build explainable AI systems, that allow for stakeholder- and domain-specific personalization. To our knowledge, barely any empirical research has been carried out focussing on the effects of interaction and iteration within AI development on XAI and the benefits and challenges that such an approach might have.

5.3. Materials and Methods

5.3.1. Action research

This study applies action research (Canterino et al., 2016) to investigate an AI development project designed to support Check, the pseudonym of a public sector organization which tackles insurance evasion. Action research aims *'to contribute both to the practical concerns of people in an immediate problematic situation and to the goals of social science'* (Rapoport, 1970). By integrating action and research, this technique addresses practical challenges arising during case studies (which may be relevant to similar cases in the same field) whilst advancing academic research (Canterino et al., 2016). The normal procedure is to set up a work

group of people in the organization and external researchers, with the dual purpose of achieving organizational project goals and contributing to academic research (Laganga, 2011). The practical-academic character of group action research can create synergy, in that researchers bring a scientific perspective, interpreting results in the light of existing literature, while organizational members contribute with practical insights gleaned from their everyday experience (Rapoport, 1970). Action research is a collaborative approach based on dynamic interaction and a variety of inputs from many actors; these must be carefully studied and selected by the researchers during their analysis of the results (Greenwood & Levin, 2007).

5.3.2. Research setting

The public sector organization called 'Check' manages compulsory insurance against work accidents and occupational diseases. Its objective is to mitigate the consequences of accidents, reduce their occurrence and safeguard workers engaged in high-risk activities. Companies are required by law to take out insurance for each employee that is commensurate to the risk associated with that person's work. This insurance provides a safety net for injured or seriously ill workers through financial benefits, supporting treatment, rehabilitation and reintegration into their professional life. Check's organizational structure is complex, consisting of an upper management layer responsible for the overall strategic direction, and various organizational sub-units or departments which implement this strategy. Each of these departments corresponds to a given function (human resources, capital management, etc.), and inter-department collaboration and coordination is actively encouraged. The various local offices carry out physical inspections on instruction from the regional level. Inspectors assess the level of risk workers face in their jobs and report accordingly to the local office in their region. Check's job is to tackle insurance evasion. Companies often declare lower risks to reduce their premiums. Check carries out scheduled inspections in companies identified as potentially at risk of evasion. In its original form,

the process to determine which these companies are is both intricate, consisting of many steps and phases, and involving numerous people from many parts of the organization.

Check decided to introduce AI to improve the process, bringing on board an external AI specialist organization, here called Innovators, to help them achieve this purpose. As researchers, we were part of the Innovators team. Checks' management resolution to augment the current inspection decision-making process through AI was underpinned by several reasons. Firstly, as mentioned, the current process is lengthy and somewhat cumbersome, and AI was expected to improve its efficiency. Secondly, AI could help to pinpoint the companies that warranted a visit. Lastly, the inspectors could be empowered in their work, and find room to address the most valuable tasks.

5.3.3. Data collection methods

The researchers collected data during meetings, interviews and a workshop, where they were either participants or observers, as well as through online participant observation and by taking field notes for 14 months, over two phases, see Table 10. The collaborative work group consisted of the researchers themselves and professionals from the Innovators team and Check members. The core project group consisted of the three authors of this paper (i.e. the researchers), a project manager representing Check, a project manager/data scientist for Innovators, and an additional data scientist working at Innovators. The project manager representing Check acted as liaison between the project team and the rest of the organization. The project manager from Innovators was the overall project leader.

The authors of this paper held different roles within the project group. The first author was mainly an observer throughout the process, while the second and third authors participated actively in the meetings and interviews, supporting Innovators' project manager by managing the

relationship with Check and asking questions. The first author asked questions indirectly through the third author.

Before starting to collect data, the researchers held two meetings with upper management representatives from the departments at Check that wanted to augment their processes and who expected to feel the impact of a new AI solution. The purpose of these meetings was to define clearly the objectives of the project, and the data to be collected, processed and implemented in the subsequent phases. The management representatives also identified key staff to help in determining the data for training the AI and strategies for its implementation.

The first phase of the data collection mostly consisted of interviews. We held seven semi-structured interviews of approximately one hour. Some of the interviews were group interviews (for a detailed overview of interview participants see Appendix A). Examples of stakeholders interview include Checks privacy expert, main innovation manager/ AI expert, a regional head inspector, and the (middle) manager of the main department involved in the this project. In total we interviewed twelve stakeholders. Earlier research has established that to be an appropriate sample size (Hennink & Kaiser, 2022). Semi-structured interviews start from a pre-determined list of topics and potential questions, but have room for the interviewees to add further information (Yin, 2017). We deliberately chose a relatively loose approach for these interviews, but made sure the same themes were consistently covered (Qu & Dumay, 2011). We approached the interviews loosely, because the topic we wanted to study is relatively new, and we were unsure about the familiarity of the various stakeholders with the topic at hand. We asked all Checks' stakeholder about their role in the organization and the difficulties they experience and their role in generating lists of companies to be inspected. Next to that, the interviews sometimes covered topics more specific to the interviewees' backgrounds and roles. For example, the privacy expert was asked much more about data privacy issues and AI, while the

innovation manager talked more about strategic innovation issues and the role of AI. We wanted the respect these different backgrounds, as every stakeholders brings their own levels of knowledge and aptitude to the topic of AI (Longo et al., 2024)

The second phase consisted mostly of meetings with various Check professionals. Within these meetings the researchers could listen in. The meetings were usually held between Innovators project team and different stakeholders from Check. On the basis of the interview material, the project team proposed an initial AI solution, holding two meetings to present the technical details of the system and gain feedback. Based on received feedback Innovators refined the initial AI solution into a draft proof of concept (PoC), and created a demonstration version for Check stakeholders to test. The researchers also had access to this tool. Innovators presented the PoC at a meeting attended by many stakeholders, who brought up several issues at this and at subsequent meetings. Innovators then started working on a new version. The second phase ended with a workshop where the project team presented the final proof of concept. A more detailed description of the interactive and iterative development approach is in section 5.3.4.

5.3.4. An interactive and iterative AI development process

We characterize the AI development approach employed in the project as 'Interactive and iterative AI development'. Innovators uses this methodology in most of their projects. This approach, while not strictly following a standardized framework, has similarities with an Agile way of working (for more info on agile see (Abrahamsson et al., 2017)). Innovators way of working is characterized by continuous collaboration with the organizations and stakeholders they work with, and flexibility to adapt to changing requirements. However, traditionally their approach also has been strongly data- and tech-oriented, meaning that the availability of any kind of data to work with was usually the starting point of many AI related projects.

Table 10: Data collection methods and use of data in analysis

Data collection methods	
Data type	*Use in analysis*
Phase I	
Primary data sources	
Presence and participation in the project kick-off meeting (duration ±45 minutes)	Enriched our understanding of the stakeholders involved and the organizational way of working
Seven semi-structured interviews (duration ±1 hour each)	Contributed to and enriched our understanding of the stakeholders' different perspectives and intentions
Presence at two meetings where the results of the initial analysis were presented (duration ±90 minutes each)	Contributed to and enriched our understanding of explainability wishes and demands
Phase II	
Primary data sources	
Presence and participation in two internal project team meetings (approximate duration 90 minutes each)	Provided insight into different ways of approaching AI development processes
Presence at five meetings with Check (approximate duration 1 hour each)	Provided insight and enriched our understanding of explainability wishes and demands, and the impact of interactivity and iterations
One semi-structured interview (approximation duration 1 hour)	Provided insight into different ways of approaching AI development processes
Secondary data sources	
Two slide presentations of the intermediate and final PoCs	Provided insight into the technical perspective on AI solutions and choices within the development
Access to PoC online demo	Provided insight into the technical perspective on AI solutions and on how the organizational stakeholders' demands were translated

We as researchers brought in additional elements to this already interactive and iterative approach. We emphasized the need for a comprehensive understanding of the stakeholder needs and the organization wherein the AI solution eventually would operate. Hence,

we pushed for additional stakeholder engagement, and exploring not only technical considerations, but also organizational and social implications of the AI. We advocated for stakeholder interviews, workplace observations, and collaborative meetings to uncover both explicit and implicit needs.

5.3.5. Data analysis

We analysed data gathered from the recording, transcription and analysis of the interviews and meetings, as well as from all the documentation we collected. After completing the interviews, the first author coded the interviews using the qualitative data analysis software ATLAS.ti. Additionally, the researchers met regularly, sharing their observations and interpretations, and, in some meetings, the coding. The first author did most of the coding, which involved reading the interview transcripts and notes made during the interviews whilst simultaneously highlighting key statements and applying open coding, meaning that the coding followed the data. This method helped us identify emergent themes in the data (Williams & Moser, 2019). Already in the earliest round of coding we noticed that the various stakeholders that we spoke to had very different expectations of AI and different concerns about what AI is and what it would change for them. This closely resembled comments from the literature regarding the different backgrounds and aptitudes of stakeholders (Longo et al., 2024; Meske et al., 2022). Connecting this to explainability we noticed that the interactive and iterative AI development approach provided the opportunity to engage with the various stakeholders and really provide them with 'personalization'. As such, we realized that the chosen approach could provide various benefits. However, at the same time we noticed in our data that many 'negative' points were brought up as well. For example, the difficulty that developers had throughout the process, and the time they had to invest. Seeing these different sides of this approach led us to frame the data as being divided into challenges and benefits. The presentation of our findings loosely resembles Hüllmann et al. (2024) in the sense that the challenges and benefits are often two sides of the same coin. The particular presented

challenges also has an upside in terms of interactivity and interaction for explainability.

5.4. Results

The results of our study are presented along the two main categories, the challenges and the benefits of interactive and iterative development approach for explainable AI models. Table 11 provides an overview of these challenges and benefits.

Table 11: Challenges and benefits of the iterative and interactive development approach

Challenges	**Benefits**
Conflicting explainability demands	Personalization of explanations
A lack of the appropriate skills among developers	Learning processes among developers
Key stakeholders are hard to commit	Selective engagement of critical stakeholders
Magnitude of time-investment	Organizational learning processes

5.4.1. Challenges

Conflicting explainability demands
We found the challenge of conflicting explainability demands. This has to do with the fact that there are differences in the type of explanation and the form of explanations that different stakeholders prefer. When interacting with stakeholders simultaneously, this can hamper understanding in both.

A first source that draws out differences in explainability demands is the role that stakeholders have within their organization. For example, within our case, middle managers wanted 'the model' explained to them. At the same time, they were also interested in the overarching

organizational implications of developing and using the AI model. In other words, what kind of organizational adjustments or policies are necessary to successfully implement AI. Additionally, they wanted to understand the organizational implications of the model, and the processes that would be impacted because of the models' development and subsequent implementation and use. We observed a contrast between the demands of middle managers and the demands of other stakeholders, like end-users. For example, a representative for Checks inspectors (the intended end-users), was interested in the model. However, this inspector was even more interested in the models' projected impact on the daily work of inspectors, and if these would be able to understand and trust the reasoning behind the models recommendations. Hence, their explainability wishes seemed to be mostly related to general reasoning behind the model, and the impact of their own daily work and role.

A second reason for conflicting explainability demands is the existence of different knowledge levels among stakeholders. The project manager highlighted: '*One of the complexities of doing machine learning is that you really need to work with a lot of different people*' [Innovators project manager]. The project manager talks about machine learning (a type of AI) because the team choose to develop a machine learning model to support the organization. He mentioned that the complexity of doing AI has to do with different knowledge levels among stakeholders. This disparity can be exemplified by the middle manager that was mentioned earlier, and a different stakeholder: Checks innovation manager. The middle managers' knowledge level about AI was minimal. Consequently, she had to be informed about the many forms of AI, its basic working and the benefits it could provide. This was unlike the innovation manager, who was well versed in the world of AI and had a much higher level of AI knowledge. Consequently, a tension existed between the approach the team had to take to cater to both of these stakeholders with very different knowledge levels. The example shows how the project team had to adjust how they talked about AI depending on who they were talking to. The project team

learned that it was necessary to cover the ground in detail with some impacted stakeholders, while with others this was less necessary. The team found that it was easier to reckon with the stakeholders' different skill levels and learning needs when on occasion they interacted with these stakeholders separately.

A lack of the appropriate skills among developers

We observed that a lack of the appropriate skills among developers can be a challenge. This became apparent throughout the entire interactive development process, and has also been mentioned in informal encounters within the project team. The lack of the appropriate skills among developers is partly due to the fact that AI development with its emphasis on interaction and iteration, is not something the developers are used to:

> *"What I've done more and I've seen done more is. As soon as we have an idea and some data to pursue the idea, we just go straight into coding or implementation or read the data and apply some algorithms"* [Innovators project manager].

Contrastingly, the more traditional process that these developers are used to is much more technically orientated, meaning that it is about getting data as quickly as possible and finding the right technical approach. Often this means that after obtaining the data, the developer team would start with an exploratory data analysis to understand the data they are working with. However, the approach followed within the case asks for a set of different skills. These skills differ from the traditional technical explanation skills and methods. Two skills can be highlighted. Firstly, the skill of questioning or to know how to ask 'the right questions'. The developers initially struggled to get out Checks' stakeholders needs and views on their own roles. However, as the process progressed questioning proved fundamental. Essentially, the developers' questioning skills nudged Checks' stakeholders to think aloud. The importance of

questioning is a novel finding in that it shifts the attention from technical in-depth knowledge to the communication and social skills necessary for explainability. Secondly, the ability to educate has proven a critical skill that developers sometimes lacked. During the many interactions, it was often necessary to educate stakeholders about something or to explain something AI related to them: *"A lot of times, so there is a really a lot of education to be done "* [Innovators project manager]. The importance of this ability highlights the importance of active explaining and educating when it comes to explainability of AI and particular models.

Key stakeholders are hard to commit

During the development process, it was occasionally necessary to engage and interact with specific stakeholders. However, the commitment of key stakeholders sometimes proved to be a challenge. The workshop and interviews uncovered two primary reasons the involvement of required stakeholders was sometimes hard. First, some stakeholders seemed unwilling to be involved in the interactive and iterative development process. A notable example is a privacy expert involved in the project. The specialist wanted to discuss general topics and his current role, but loathed to answer specific questions about the risk monitoring process, its impact, and the potential role of AI. When the project team persisted, his answers were concise: "*As for the rest, I don't know the details of this contract, or if there even is one*" [Checks' Privacy expert]. Second, it also seems some stakeholders do not understand why they are needed. Hence, getting their involvement is difficult. This may also have to do with the level of knowledge they have regarding AI, as has been mentioned before. Essentially, when necessary, some stakeholders apparently wanted to be more involved, while we found others uninterested in having a say beyond their own current responsibilities.

Magnitude of time-investment

A development process characterized by constant interaction and iteration is time intensive. The project team interacted with organizational

stakeholders in many different ways and at many different times. The idea behind that investment is that this will pay for itself later on, as noted by Innovators project manager: "*Investing a little bit of time now can result in having good tools tomorrow*" [Innovators project manager]. However, it is not certain that this considerable time investment will provide results later on. There is no certainty that the time investment will lead to 'good' and explainable AI tools that are interpretable and hence accepted by organizational stakeholders. Thus, there is a risk associated with the time investment, as it might not pay off. Additionally, this time-investment, especially time associated with the 'social' tasks surrounding explainability, takes time away from the more technically oriented tasks, that also might contribute to higher quality models that are technically well explainable. An illustrative example is related to the data. As referred to earlier, the developer team's main goal is usually getting the data as quickly as possible. In this instance, getting the data proved to be an extremely lengthy process. This seemed partly to be related to the approach, as the developer team first took a considerable amount interacting with the stakeholders, before even asking about potential data.

5.4.2. Benefits

Personalization of explanations

The personalization of explanations represents a strong benefit of the approach. We observed this throughout the entire process. The utilisation of interactivity and engagement with a multitude of stakeholders throughout the process enabled the developer team to gain a understanding of the specific needs and characteristics of each individual stakeholder: "*It's useful to get into their world right, and understand more or less how they think, how they act, and how they work*" [Innovators project manager]. Here, the example mentioned earlier that highlighted differences in knowledge levels between the innovation manager and the middle manager is also exemplary. Explaining the general workings

of AI was totally unnecessary for the innovation manager, due to his level of knowledge and background. The approach taken helped the developer team to deal with these differences. A key characteristic of the process that strongly contributes to dealing with these differences is fluidity. Fluidity refers to the fact that the process was not set in stone, that different levels of stakeholder involvement was possible, and that Check stakeholders were able to 'move in and out' of the development process. Throughout the process the developer team understood that fluidity was key to cater to the various explainability demands. To provide fluidity within the process, it also meant that developers had to accept working with stakeholders in different assemblies. For developers, it also required flexibility in their attitude and approach. Consequently, they were able to personalize the explanation for the middle managers towards explanations about the impact on their departmental strategy, for the privacy specialist towards explanations about the data behind the algorithms, and for the end-users explanations focused on their working processes and the general workings of the model.

Learning processes among developers

Throughout the development learning processes occurred within the developer team. These learning process were related to their own skills, but also to level of knowledge the developer team had about Check and its main stakeholders. Firstly, the developers, throughout continuous engagement with Check's stakeholders became increasingly skilled at identifying the most appropriate terminology for different stakeholders and at conveying information in a manner that was accessible to these stakeholders. Second, the developer team learned about 'the right problem to solve': "*It's very easy to solve the wrong problem, because you have some data. And then you solve the problem that is easier to solve based on that data. But it's not a given that this is a problem that the company wants to solve*" [Innovators project manager]. As stated, in the more traditional 'tech-heavy' approach that Innovators often followed. the first step would be to request the data from the organization, which

would then be shared with them. However, there is a risk of solving the wrong problem. Frequently, organizations like Check are not aware about the issue they want to solve with AI. Hence, by immediately being asked to share data, organizations often share irrelevant data for the problem they want to solve. The interactive and iterative approach, with more emphasis on getting to know the main stakeholders, their needs, and their level of knowledge, helped developers to get a clearer picture of the right problem to solve. Third, the developer team understood who to turn to with specific issues and questions:

> *"We have the right data, we have the right problem in mind. We know the right people to talk to. If we have architectural problems, infrastructure problem, privacy problems. We have a dictionary of people to go to."* [Innovators project manager].

The statement highlights how the developer team - throughout the process.- gained knowledge about who to turn to with specific issues and questions. The organization become interpretable to the developer team. They expected that this knowledge would help to ease the process later on, when an actual AI solution would be build and implemented.

Selective engagement of critical stakeholders

The selective engagement of critical stakeholders proved to be a benefit. The flexibility ingrained in the development process gave the opportunity for the team to meet with stakeholders in varying configurations. Likewise, the organizational stakeholders had freedom to move in and out of the process. The benefits that this provided can be best illustrated with the description of a situation that occurred during a presentation by Innovators developer team. Innovators project manager and head developer presented the first results in a meeting with Check stakeholders, which included the main middle managers. However, because of the function of the particular meeting, these middle managers invited a statistician. In the particular meeting the Innovators presented the first PoC after

having spoken to multiple stakeholders within Check. This PoC had a feature that gave users the choice between different types of algorithms. The results of the PoC could vary depending on the algorithm selected. A middle manager and mostly the statistician both criticized this point, because they felt that imposing these choices on the end users would be unhelpful. In response to the comments, the developers changed the PoC substantially, taking a different direction regarding data segmentation, while also reducing operational complexity for the end users. In the new version, they were no longer responsible for these decisions, which were instead delegated to other internal stakeholders with more centralized roles. The point we want to illustrate is that the development approach gave room for the statistician to be present. However, another finding emerged. The example also illustrates that individuals who can speak both the language of AI and the language of the application context play a critical role in the development of AI tools and in making AI explainable throughout the organization and making the organization explainable to the developers. In this example, the statistician played a crucial role, because this person was knowledgeable about AI – the world of the developers from Innovators – but also about the domain in which AI would be used – the world of Check stakeholders.

Organizational learning processes

As an organization Check strongly benefited from learning processes. Innovators Project manager mentioned: *"I think in this sense, they are also growing in terms of knowledge and I think this is proving to be useful."* The constant interaction between the developer team and a multitude of Check stakeholders stimulated and compelled them to start learning about AI. The project manager mentioned that these learning processes that occurred, were helpful later in the development process, because the stakeholders understood much better what AI was about, what they wanted, and what their organization was still lacking. These learning processes started already early on and took place in various levels. First on the level of stakeholders, in this case middle-managers. Within the

first official project meeting, the head of the Insurance Evasion Risk Office (a sub-department run by a middle manager) was keen to understand more about AI:

> *"I want to learn about machine learning and get familiar with the tools we plan to introduce at Check. For me, it makes sense to find out as much as I can about their features and potential"* [Checks' middle manager]

This appetite for learning, here displayed by a middle manager was important. It helped later in the development process, when this middle manager was able to provide substantial comments about the proposed AI solution, instead of only superficial comments. These learning processes helped Check's middle managers understand and manage AI's impact on the process AI was supposed to support (tackling risk evasion). These learning processes relate to XAI in that they make AI more explainable to different types of stakeholders, and expose the broader nature of AI, as well as the complexity of developing an AI solution for a large public organization. Secondly, learning took place in terms of data readiness and data understanding. This was illustrated when the middle manager responsible for digital services within Risk mentioned: "*It is a stimulus to the quality of the data that afterwards returns benefits elsewhere.*" [head of the Digital Services Office]. He referred to the fact that the interactive and iterative approach stimulated Risk to enhance the quality of their data. Because of increasing knowledge throughout the process, Risk understood that the current quality and organization of their data was not yet sufficient to support the development and use of AI tools. These learning processes were facilitated and fuelled by the design of the development process. Check's stakeholders shared their appreciation for the development approach provided: '*We also thank you specifically for your openness to other solutions*' [head of the Insurance Evasion Risk Office]. The middle manager mentioned this, because throughout the development process, the technical AI solution changed substantially

after operational staff learned that the AI solutions proposed in that moment were not appropriate for the organization.

5.5. Discussion

After presenting the benefits and challenges, in this part, we will discuss and summarize the main findings, discuss the main contributions to the literature, and give implications and recommendations for practitioners. We will conclude by discussing the main limitations and future avenues of research.

5.5.1 Summary of findings

Studying the impact of an interactive and iterative development process on the explainability of AI models reveals several key findings. First and foremost, our findings show that explainability is a process of reciprocal learning. In this process, there is no single lead player (e.g. a developer) who applies technical explainability techniques to a particular model. Instead, the process is about ongoing learning between and among actors. The developers learn about the organization, while the organizational stakeholders learn about AI. These learning processes are a direct consequence of various forms of involvement in the AI development process. And because of these learning processes, AI becomes more 'explainable' to the stakeholders within the organization that are involved or expected to be affected by implementation of the AI model. This novel understanding adds to the existing XAI and information systems literature that has shifted its attention to the various XAI stakeholders with their specific explainability demands (e.g. Langer et al., 2021; Longo et al., 2024; Meske et al., 2022; Tomsett et al., 2018) and literature that discusses the potential of interaction within the AI development process as a source for explainability (e.g. Bhatt et al., 2020; de Bruijn et al., 2021; Longo et al., 2024). Our work confirms that the AI development process can be a source of explainability. However, not only in the sense of explaining 'the inner

workings' of a model. Instead, the AI development process contributes to a 'broader' form of explainability in which organizational implications and the impact on the roles of involved stakeholders becomes visible.

Secondly, we found that the demands of explainability and the desired types of explanations are linked to the stakeholders' current roles within the organization. This finding is partially in line with previous literature stating that different types of explanations are needed for different types of stakeholders (Langer et al., 2021; Meske et al., 2022; Preece et al., 2018). However, previous work has not explicitly focused only on internal organizational stakeholders. Based on our findings, we propose a classification of internal stakeholders into six categories. The different types of stakeholders and preferred types of explanation are given in Table 3. The table has a column that refers to the type of explanation. This column refers to what we have seen these particular stakeholders to be interested within the case. For example, overly technical details are not preferred by non-specialists (like middle management). However, context-aware specialist (that span the boundary between developers and other organizational stakeholders) were much more interested in explanations regarding the underlying data and specific instances of input-output relationships.

Table 12: Overview of internal stakeholders and types of explanations

Stakeholder type	***Type of explanation***
Middle management	Explanations about the impact on their departmental strategy (e.g. explanations about impact on the employees' daily work).
Cross-discipline specialists	Explanations connected to their specific area of expertise (e.g. does the AI solution fit privacy requirements? what must I do to facilitate the AI solution?)
Context-aware specialists	Explanations about underlying data (e.g. is the data integrated correctly? does it fit daily reality?)
Operational end users	Explanations about the impact on daily work (e.g. does it make the inspectors' lives easier? could it impact their work?)

Thirdly, our findings indicate that two groups of stakeholders in the learning process are fundamental for these processes to occur, these being the context-aware specialists and developers. Context-aware specialists are in the position to understand the nuances and choices behind the algorithms on a deep technological level, and also understand the context and nuances of the area where these algorithms will be used. Hence, these stakeholders act as so-called boundary spanners (Williams, 2013), linking the developers and the other organizational stakeholders and translating the insights from within the organization to the external developers and the technical expertise of the developers to the domain of the organization . Developers are critical because their position essentially leads them to by the main instigator of the organizational learning processes. Because of their position, they 'set the scene' and conditions for learning. However, their critical position to facilitate these learning processes is also dependent on their own wants, needs, and skillset. This brings is to the last finding.

Finally, our findings indicate the need for developers to be endowed with new skills. Alongside their technical expertise – which in terms of XAI is often model-centric - developers are asked to acquire skills traditionally known as 'soft skills' such as questioning and educating. The interactive

and iterative AI development process demands that the role of the developer evolves.

5.5.2. Contributions to research

Our findings contribute to the literature on explainability in various ways. Firstly, as discussed, much of the explainability literature draws attention to the fact that different stakeholders want different explanations (Barredo Arrieta et al., 2020; Belle & Papantonis, 2021; Bhatt et al., 2020; Langer et al., 2021; Meske et al., 2022; Preece et al., 2018; Tomsett et al., 2018). However, so far the literature has not focused on internal organizational stakeholders. Our work provides a typology of different types of organizational stakeholders within a public sector organization and the various types of explainability these stakeholders prefer. Different stakeholders may want/need different explanations, ranging from the technical aspects of AI and the more classic 'model-centric' explanations to organizational consequences of implementing AI models. As such, a redefinition of what XAI is actually about. In line with the work by Langer et al. (2021), we found that, in reality, the stakeholders' needs may be even more subjective, and thus stakeholder categories even more fine-grained. Hence, we concur with the need for personalized approaches and explanations. Prototypes and stakeholder categories must be taken into account, but in reality, the demands for explainability can go even beyond categorizations.

Secondly, our contribution to the literature lies in a reconceptualization of the *what* of XAI. Much literature discusses the need to explain the system (Tomsett et al., 2018), models (Barredo Arrieta et al., 2020) or specific instances of the model (Belle & Papantonis, 2021; Bhatt et al., 2020). Our findings confirm the importance of explainability that is model-centric. However, somewhat counterintuitively we also found that in the messy reality of AI development explainability is even broader in scope. Hence, involved stakeholders do not only want the models explained to them. Instead the training data, the choices during the development process,

and the implications for their daily work are also part of the demanded explanations. These are often heavily dependent on their own background and knowledge levels, as noted by (Longo et al., 2024). Consequently, the mentioned aspects could also be framed as part of XAI. Thus, our work is aligned with the work by de Bruijn et al. (2021) who discussed the notion of explainable processes. However, as said, our findings suggest that there is an even broader and more detailed notion of what should be explained. This is necessary as the impact of the AI application is ultimately dependent upon the stakeholders understanding of it.

Lastly our work adds to the literature that discusses the need to interact with stakeholders for explainability (Bhatt et al., 2020; de Bruijn et al., 2021; Longo et al., 2024). Our work gives a more detailed and empirical understanding of how explainability can be achieved through the AI development process and highlights the main benefits and challenges. The findings suggest that interaction for explainability is often messy because of contrasting wishes, knowledge levels and needs. At the same time interaction is a catalyst for organizational learning processes. These processes give organizational stakeholders a greater knowledge base to understand AI. As such, model-centric explainability is not necessary increased. However, these processes help them understand what the particular AI application entails within their organizational context and for their specific role. Also,, our work points in the direction of critical actors (i.e. developers and context-aware specialists) and critical skills (e.g. communication skills like questioning) that are necessary to ensure successful interaction.

5.5.3. Implications and recommendations for practice

These findings also have implications for practitioners. Firstly, we encourage practitioners to explicitly make space for learning in the development process of public AI tools. Upper and middle managers can do so in various ways, for example by 1) giving their employees the time and resources to be involved in the process and, 2) giving them

the freedom to choose their level of involvement in the process. The probable outcome is a more explainable AI application. Stakeholders who understand the application will have more trust in it, to the benefit of both themselves and their organization.

Secondly, we encourage practitioners, and especially developers, to develop a set of additional skills on top of more traditional and technical skillsets. Naturally, developers and data scientists often come from a strong technical background. However, there is sometimes less emphasis on the organizational and social aspects of developing AI solutions. We advocate for skills such as questioning, mapping stakeholders and their intentions, and presenting complex topics in easy ways, all skills that can maximize the value derived from interactions between developers and internal organizational stakeholders, and so contribute to more broadly accepted and explainable AI solutions. However, additional skills require time and resources, and we encourage upper and middle managers to provide both and even see it as an investment, albeit one where the benefits are reaped later.

Lastly, we encourage public sector leaders to give their backing to organizational stakeholders who know how to move between domains. AI applications, in particular, bring into play many disciplines. Boundary spanners seem critical because they can make AI solutions explainable in various domains and are able to transfer sometimes difficult concepts across domains. Boundary spanning individuals should have an assured place in highly specialized public organizations where structures, roles and responsibilities are rigid.

5.5.4. Limitations and future research

Our study is not without limitations. Firstly, the generalizability of our findings might be limited. Our study is based on a single case in the public sector within a particular cultural context, and our findings may not be representative of other types of public sector organizations or

in different cultural contexts. Therefore, we encourage other scholars to undertake additional empirical research into this novel conception of explainability and the impact of these types of development processes, for example through additional (multiple) case studies in different countries. The second limitation concerns the data analysis. While we triangulated data, worked in a three-person team and coded the data following a well-structured plan, the findings might partially reflect the authors' subjective bias. Thirdly, our research covered the early phases of the AI application, hence only a specific phase of the AI lifecycle, without studying its development and use from beginning to end. To gain a fuller picture, we would encourage scholars to study the entire process, and to include all the application's end users, as in our case, it was not possible to speak directly to the inspectors who would be acting under the new system. Finally, our study only touched upon the impact of interactivity and an iterative approach on explainability, whilst observing that the many facets of AI are closely interrelated and therefore not easily untangled. We encourage others to research the impact of this type of development processes on issues like accountability and sustainability, and how these matters interconnect. A holistic and multi-faceted understanding of helpful, ethical and trusted AI systems is a must to leverage on the great potential of AI applications in the public sector.

5.6. Conclusions

Our study aims to answer the questions of what the benefits and challenges of an interactive and iterative development approach for explainability of AI models are, and how do the benefits and challenges occur? Our work builds on earlier information systems studies and answers the call for more empirical research. Hence, we studied an AI development project for an Italian public organization responsible for countering mandatory insurance payment evasion in the business sector, which intended to contribute to a key decision-making process

to determine which companies it should inspect. We were especially interested in how an interactive and iterative approach contributes to AI explainability, as suggested in the literature (Bhatt et al., 2020; de Bruijn et al., 2021; Longo et al., 2024) .

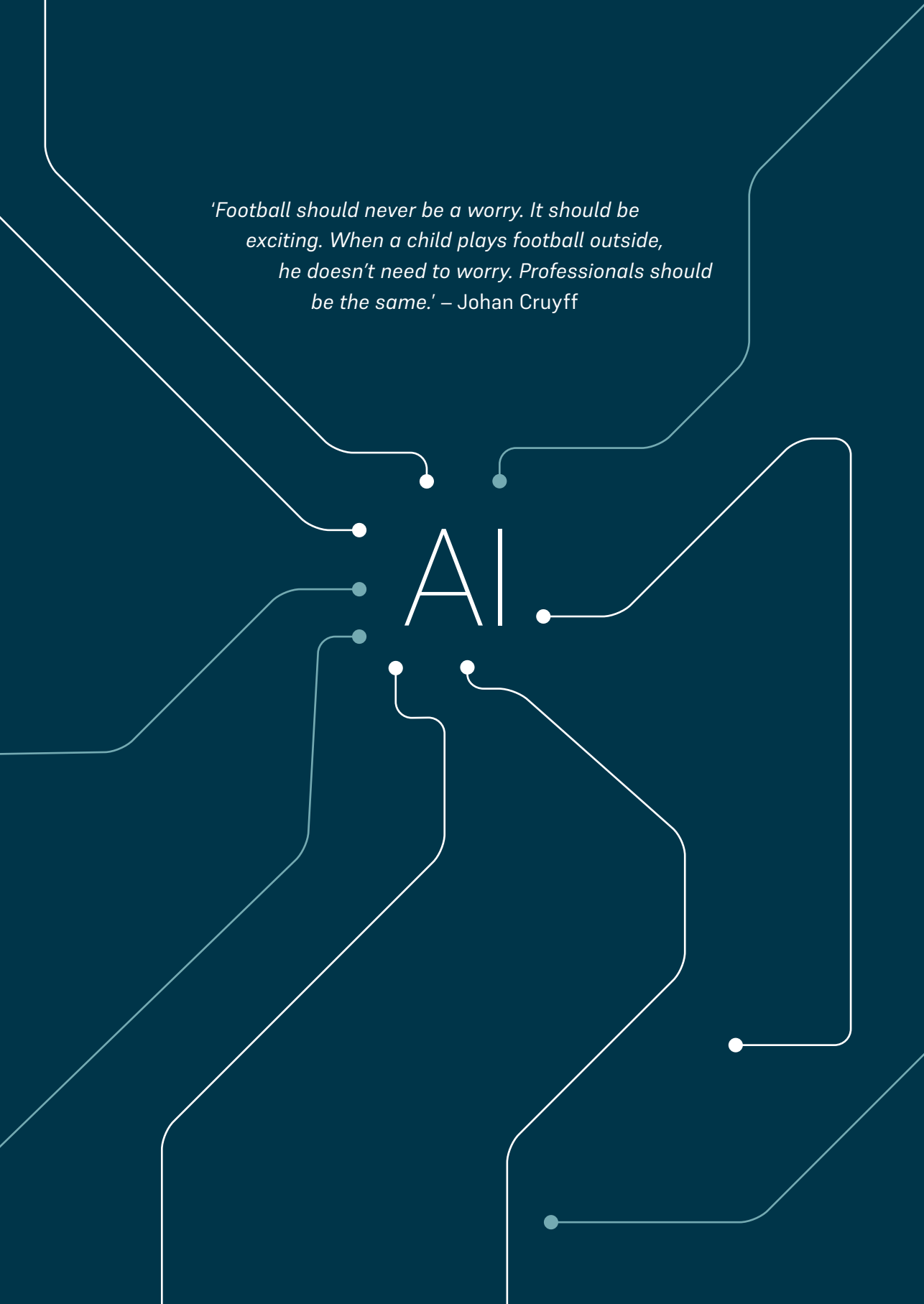
We found that an interactive and iterative development approach contributes to explainability in multiple ways: through instigating organizational learning processes, and through the personalization of explanations. However, the interactive explainability process also has many challenges, including involving the required people at the right moment, the strong dependence of developers and their novel skillset, and conflicting explainability demands.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used DeepL in order to improve readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

5

'Football should never be a worry. It should be exciting. When a child plays football outside, he doesn't need to worry. Professionals should be the same.' – Johan Cruyff



6

The Impact of Generative AI on Inspectorates: An Empirical Exploration among Professionals

This chapter is based on a journal paper originally written in Dutch. For the purpose of this dissertation the paper has been translated to English. The paper is published as: Krimpen, F. van, Voort, H. van der, & Bruijn, H. de (2025). De impact van generatieve AI op toezichthouders: Een empirische verkenning onder professionals. Tijdschrift voor Toezicht, 16(2-3), 44-55. https://doi.org/10.5553/TvT/187987052025016002002.

Abstract

Generative AI (GenAI), according to many, will have a major effect on professionals in the public sector. This means that professionals within inspectorates will also be affected. These professionals perform an important function and often operate relatively autonomously and on the basis of exclusive and also often implicit knowledge. GenAI is relatively new, and thus empirical research on the impact of GenAI on the daily work of professionals is scarce. This study aims to understand the impact of GenAI on professionals within inspectorates by surveying the professionals themselves. We found that professionals are positive about the potential of GenAI for their work. We also see that the relationship between professional and GenAI becomes not one-sided, but multi-sided: GenAI becomes a sparring partner. A prerequisite for a good relationship between professional and GenAI is enabling learning processes for professionals, by giving them space to experiment.

Keywords
Generative AI, ChatGPT, professionals, autonomy, learning processes

6.1. Introduction

Generative AI (GenAI) is a new step in the development of AI. GenAI systems can mimic human creativity (Ritala et al., 2023) and increase the productivity and quality of knowledge work (Dell'Acqua et al., 2023). The now best-known application is ChatGPT - a publicly available tool based on a Large Language Model (LLM), which is widely used. Generative AI systems are used for tasks as diverse as writing business plans, legal documents and software codes. Also summarising texts, translations or writing e-mails and other messages can easily be done by GenAI tools. There is research showing that GenAI can outperform human workers for certain types of tasks (Gilardi et al., 2023).

Not surprisingly, professionals within the public sector want to use these tools or are already doing so. A recent survey by Bright et.al (Bright et al., 2024), who surveyed 938 public sector professionals in the UK (in the fields of education, healthcare, social work and emergency services), found that almost half of respondents were aware that GenAI was being used within their area of work. Indeed, 22% of respondents said they were actively using a GenAI system.

Many GenAI systems are extremely accessible, partly because they often have a free version. Public professionals can therefore easily use GenAI. But GenAI is a new and revolutionary phenomenon and this means, among other things, that there are different views on the desirability and undesirability of public professionals using Gen AI. There are several concerns, such as data privacy, transparency and quality of GenAI output. In late 2023, then-Secretary of State Alexandra van Huffelen wrote in a parliamentary letter that the use of non-contracted GenAI applications by government organizations is in principle not allowed, partly due to concerns around data privacy and the 'black box' nature of the tools (Van Huffelen, 2023) . This has certainly affected usage. Nevertheless, GenAI

is expected to change the work of public professionals. In this article, we focus on a specific group of public professionals: professionals working in inspectorates (i.e. supervisory professionals).

Not much is yet known about the use of GenAI by supervisory professionals. Recently, Dell'Acqua et. al (2023) described the impact of GenAI on certain types of tasks. While this work is not focused on supervisory professionals, it does show that GenAI can contribute to the type of work that professionals within inspectorates perform. Indeed, the tasks studied are all related to knowledge work - and supervisory professionals are also knowledge workers. Research like Dell'Acqua et al. (2023) does not take into account the certain specific characteristics of professionals. For instance, professionals possess tacit knowledge, which is difficult to objectify. The complexity of the tasks they perform often produces value conflicts - and these too cannot always be objectified. The question is how a GenAI tool relates to these non-objectifiable aspects of professional task performance.

In addition, recent work particularly describes how GenAI can impact professionals, but goes less into the views of the professionals themselves on the use of GenAI. Similarly, the Secretary of State's policy mentioned above was at the time without consulting professionals 'on-the-ground' - even though the work of these professionals can be strongly affected by these policy choices. Therefore, this research seeks to understand how (and whether) professionals within supervision use GenAI tools, and more specifically how they expect it to affect their daily work. The main question of our research is: *How do professionals within supervisors expect generative AI to influence their work?*

Our approach is based on seven semi-structured interviews with professionals within one supervisory agency and a workshop with about 60 professionals from different supervisory agencies. We identified the perspectives of these professionals on the expected impact of

generative AI on their work. To this end, we distinguished four aspects of professionalism, which we will elaborate on in the next chapter. With this focus on professionalism, we want to contribute to the growing debate on the impact of GenAI on supervisory organizations. After all, the voice of professionals is essential in policy on GenAI. They are closest to practice, and in addition they are dealing with an environment that is freely experimenting with all kinds of GenAI tools, whether they are freely available or not. We use the research findings to set an agenda for the future conversation about GenAI and supervisory professionals, and provide recommendations for future policy.

Our article is structured as follows. In section 2, we discuss the theoretical insights, which are important when talking about supervisory professionals and GenAI. We then describe our methodology, after which in section 4 we share our empirical findings. Then, in Chapter 5, we discuss these findings in light of our conceptualization of supervisory professionals. Finally, in Chapter 6, we answer the main question, and offer salient insights and opportunities for future research.

6.2. Theoretical insights

6.2.1. What is generative AI?

Generative AI is a form of artificial intelligence (AI) that has received significant attention in recent years. GenAI is a subset of machine learning (ML), which in turn is a subset of the broader AI category. GenAI is therefore part of the larger AI field. Generative AI models, trained on large amounts of data, can generate new content such as text, images, music, or video (Brynjolfsson et al., 2023). The critical role of data is something in which ML and GenAI strongly resemble each other. Both forms of AI can only exist because they have been trained on available data. However, there are also differences.

First, there is a major difference in application. ML systems can 'discriminate' by, for example, classifying email as spam or not spam and assigning a probability score to this classification. GenAI systems generate 'new' content based on the enormous amount of data on which they have been trained. This generated content is not explicitly present in the training data but is based on learned patterns and relationships. This makes GenAI inherently multifunctional and applicable to multiple types of tasks (Ministerie van Binnenlandse Zaken en Overheidsrelaties, 2024), unlike ML models. Second, many current GenAI models have a chatbot interface, making GenAI relatively accessible to professionals without technical backgrounds, in contrast to other AI applications. The well-known example is ChatGPT, a service that allows users to 'talk' with a GenAI model. Both arguments show that GenAI differs inherently from other AI applications. This supports the idea that additional research into GenAI is needed, especially when used by public sector professionals.

6.2.2. Generative AI and its impact on public professionals

Due to the rapid emergence of GenAI, there is still limited scientific literature available on the use of GenAI by public professionals, let alone professionals within regulatory agencies. However, there is a longer-established and substantial body of literature on the use of AI (the broader category under which GenAI falls) within regulatory agencies. For example, Busuioc (2022) describes how public regulators are increasingly using ML techniques. She also describes the many challenges that accompany this, such as potential discrimination, but also inadequate transparency about how AI models actually work. Yeung (2018) also extensively describes how AI is applied by regulatory agencies s and identifies a broad range of challenges, such as challenges around privacy or ensuring that those affected by algorithmic decisions can appeal against them. Such challenges, however, are more general in nature and describe less how professionals are influenced by AI.

Work that does address this includes, for example, that of Lorenz, van Erp and Meijer (Lorenz et al., 2022). They note that AI may potentially reduce the discretionary space of AI tool end-users. This may be problematic because end-users typically need discretionary space to handle ambiguous situations where rules cannot be directly applied. Although these findings offer various starting points, the question arises as to what extent these findings about algorithmic and AI systems also apply to GenAI. As described in 2.1, GenAI differs from other AI applications through its ability to create new content rather than only analyzing or categorizing existing data. GenAI can also be used by professionals without the intervention of analysts or managers. Initially, this appears to increase the autonomy of professionals. On the other hand, there is also the concern that the work of professionals may be partially taken over by GenAI, which would actually reduce their autonomy. These differences indicate that it is necessary not only to describe and analyze more general literature on AI and regulatory agencies, but also literature that focuses on GenAI.

Initial studies do not focus specifically on the public sector., but discuss GenAI's impact on knowledge work. For example, Ritala et al. (2023) show how GenAI can affect different types of knowledge work (e.g. creative work versus repetitive work). They discuss that especially simple repetitive tasks, such as summarising notes or generating reports, can be easily done by ChatGPT. Also Dell'Acqua et al. (2023) conclude, based on experiments with consultants, i.e. knowledge workers, that *large language models* (LLMs), such as ChatGPT, are very capable of increasing the quality and productivity of work. This applies to both structured and creative tasks, such as brainstorms on marketing campaigns. They also concluded that consultants who could use a *prompt overview* performed even better. A *prompt overview* is an overview, which lists possible *prompts* (types of input a user gives to the model) that allow the user to query or steer the model in different ways. The better the prompts, the higher the quality of GenAI's output (Knoth et al., 2024). Although

Dell'Acqua et al. (2023) focus on consultants within their study, their findings may also be relevant for knowledge workers in the public domain, including regulatory agencies.

Literature focusing more explicitly on the public sector has been somewhat more speculative in nature. Strauss (2023) indicates that GenAI could increase the productivity of 'office workers' in the public sector . A study by Bright et al. (2024) shows that, especially in the UK, the use of generative AI is widespread in the public sector. The study also mentions that public professionals expect GenAI to have a major impact on the time they spend on bureaucracy, dramatically changing their work. This finding seems logical, given the often repetitive and simple nature of bureaucratic tasks. These are the kinds of tasks where GenAI could potentially add a lot of value.

None of the studies mentioned explicitly address the characteristics of professionalism. Dell'Acqua et al. (2023) mentions "knowledge workers", without decomposing the term or to distinguish types of knowledge workers. They mainly focus their experiment on comparing tasks using GenAI and without using GenAI. Only a few studies in the field of generative AI, knowledge work and the public sector go deeper into the characteristics of knowledge workers. Brynjolfsson et al. (2023) mention tacit knowledge in relation to GenAI. Tacit knowledge refers to the knowledge and skills that individuals possess but cannot explicitly express. It is often intuitive and non-verbal, acquired through personal experiences, observations and practice over time (Polanyi, 1966) . Brynjolfsson et al. (2023) hypothesize that GenAI chatbots are able to partially codify, and thus make explicit, the tacit knowledge of highly skilled and experienced workers. As a result, less experienced workers are able to use and leverage this previously tacit knowledge faster than before.

This theoretical notion calls for more research into the relationship between generative AI and tacit knowledge. Tacit knowledge is seen as an important characteristic of public professionals, but there is more to it. A more detailed understanding of 'professionalism' in the context of supervision will help understand the impact of GenAI. A further conceptualization of supervisory professionals is therefore required.

6.2.3. Conceptualization of supervisory professionals

Professionalism is a much studied topic. However, there is no agreement on how to define professionalism, let alone public professionalism (de Bruijn, 2012; Noordegraaf, 2007). We conceptualize public professionals using common concepts from the literature, focused on supervisory professionals. This conceptualization provides guidance for the rest of the research and later helps us interpret the results. We conceptualise professionals within supervisors as follows:

1. Supervisory professionals often rely heavily on 'tacit' knowledge - on their professional intuition.
2. Supervisory professionals have a high degree of autonomy and discretion to make decisions about and in their daily work.
3. Supervisory professionals must deal with conflicting perspectives and values.
4. Supervisory professionals carry out their work in relation to inspectees.

First, supervisory professionals rely on '*tacit knowledge*', or tacit knowledge (de Bruijn, 2012) Howells (1996) notes that tacit knowledge is often acquired through action, through experiences, observations and 'learning by doing'. This knowledge is often intuitive and non-verbal, and is accumulated over time. Tacit knowledge can be contrasted with 'explicit knowledge'. Explicit knowledge can be recorded and encoded in language or images (Nonaka & von Krogh, 2009) . Like tacit knowledge, explicit knowledge is also a necessary source of knowledge for public

professionals in their work. Both types of knowledge form a continuum, and certain tacit knowledge can be codified over time and thus become explicit knowledge. At the same time, there is also a tension between the two types of knowledge. Indeed, making tacit knowledge explicit is not always possible, and can potentially corrupt the essence of tacit knowledge. The question, of course, is what impact GenAI will have on these forms of knowledge. Will GenAI reduce the need for tacit knowledge? Or, on the contrary, increase it?

The second characteristic is that public professionals have a high degree of autonomy and discretion to make decisions in and about their daily work. The complexity of the tasks performed by public professionals limits the scope for top-down steering, even in highly regulated environments. Lipsky (1980) talks about *street-level bureaucrats*: public professionals who, in their daily work, have direct contact with citizens and have to make decisions that can strongly affect the lives of these citizens. These public professionals often work in environments with limited resources and where policies have to be adapted to specific situations in practice (discretion). Inspections can also be characterized as street-level organizations, in which the professionals involved must have autonomy and discretion. The inspector, for example, must be able to be flexible during inspections or other work to adapt his choices to the situation in the moment. Autonomy and discretion imply that professionals must also be able to be creative - the street-level bureaucrat is also the creative bureaucrat: to deal with the variety of situations, these professionals will have to be creative. Also with this characteristic the question is what GenAI will mean. Will GenAI curb autonomy? Will it affect professionals' creativity? Or does discretion shift from these professionals to other professionals within the organization, as the work of Bovens and Zouridis (2002) suggests?

The third characteristic within our conceptualization is about dealing with conflicting perspectives and values. Examples of such perspectives or

values are efficiency, equal treatment, application of rules, treatment. In the situation where an inspector has to make decisions, this characteristic of public professionals is strongly expressed. Does the inspector opt for efficiency and some flexibility regarding the application of rules, and does the supervised organisation receive a warning, or is the inspector strict, and does he put the organisation under aggravated supervision, with the processes and work involved? Some tasks of supervisory professionals are repetitive, and the trade-off between perspectives and values may lend itself to standardisation. But in other situations, this is not the case. Again, the question arises what the impact of GenAI will be on dealing with conflicting perspectives. Do we actually know whether GenAI applications can handle conflicting perspectives?

As a final characteristic, we mention that supervisory professionals carry out their work in relation to inspectees. Different types of inspectees will exhibit different types of behaviour. Well-known distinctions in the supervision literature are those between well-intentioned and malicious inspectees or between inspectees who do or do not have the knowledge to act in a norm-compliant manner (De Bruijn & Ten Heuvelhof, 2019) . Inspectees can also exhibit all kinds of strategic behaviour or '*game playing*'. Think, for example, of the '*managerial game*' (the inspector is confronted with ever-changing representatives of the inspector) or that of 'strategic confessions' (the inspectee confesses something the inspector has known all along and thereby suggests that he has made an important concession). Supervisory professionals will have to recognize and respond to these behaviours - and thus the behaviour of inspectees helps determine the professionalism of inspectors. Again, the question is: can GenAI help to map and better recognize these behaviours of inspectees? And can GenAI help develop strategies to deal with it?

This conceptualization of supervisory professionals helps us to organise and interpret our empirical findings. Also, to explore in detail how GenAI will affect professionalism. To date, very little is known about

this question. For example, as mentioned, recent work suggests that GenAI may be able to make tacit knowledge less tacit (Brynjolfsson et al., 2023) . According to this logic, the unique role of the public professional may be constrained. After all, if tacit knowledge can be made explicit through a GenAI tool, then the tacit knowledge of the professional becomes less important. Others might argue that this is not the case, and that professionalism cannot be captured by GenAI. Without passing judgement on this now, our conceptualization and what is known about GenAI to date indicates *that GenAI will change the professionalism of public professionals within regulatory agencies.*

With this study, we seek to contribute to the debate on GenAI and professionalism. We provide insight into the perspectives of regulatory professionals on the impact that generative AI will have on their professional work. This bottom-up and empirical perspective can inform future policy by regulators regarding the use of GenAI for regulatory professionals.

6.3. Methodology

This research is exploratory in nature, given the relatively new nature of GenAI and the lack of research on its impact on professionalism. This research uses two research methods, semi-structured interviews with eight employees of the Health & Youth Inspectorate and a workshop with sixty participants entitled '*Generative AI for Supervisors*' at the Supervision Festival 2024.

6.3.1 Data collection

Semi-structured interviews are interviews with a predetermined ideal sequence, but also with a lot of flexibility (Longhurst, 2010) . This type of interviews allows the interviewee to address topics that they see as important. This type of interview allows for exploration and follow-up

based on the participants' answers. For the interview, the interviewer determines main themes, a structure and a number of related questions.

Our interview protocol focused on three themes: the role of the interviewee, how generative AI can potentially contribute to their daily work, and the conditions that are important to make GenAI use successful within the organization. Frequently asked questions were: What is your role? Do you use any form of Generative AI in your daily work? How can Generative AI help you improve the quality of your work; What do you need to be able to handle Generative AI well? Additionally, the interviews addressed what respondents understood by Generative AI when necessary. We deliberately avoided being too directive in this regard, so when respondents' answers deviated from what GenAI actually is, we consciously chose not to correct them extensively. We did frequently use ChatGPT as an example, since we expected interviewees would be familiar with it.

The semi-structured interviews took place between the 1st of May and the 1st of July. We conducted a total of seven interviews, with one of the interviews being a double interview with two interviewees at the same time. The interviewees had different roles. For example, we spoke with inspectors, but also with a team leader, department heads, and a consultant. These different backgrounds allowed for different emphases within the interviews. Table 13 provides an overview of the interviewees.

Table 13: Overview of interviewees

Code	Role within supervisor
I1	Department head
I2	Inspector
I3	Inspector
I4	Advisor
I5	Department head
I6	Team leader
I7	Research assistant
I8	Data scientist

Workshops as a research method are a relatively little studied phenomenon. We rely here mainly on work by Ørngreen and Levinsen (2017). Using a workshop as a research method usually aims to produce reliable and valid data. It often involves forward-looking processes, such as organizational change. Workshops are often used for studies that are unpredictable and related to practice. Workshops as a research method often have inherent contradictions in them, as conflicting goals and roles may be present. Namely, for the researchers, the workshop serves to generate data, but at the same time it also prioritises the needs of the participants and aims to give them something as well (Darsø, 2001).

The workshop '*Generative AI for Supervisors*' took place at the Supervision Festival 2024 on the 18th of June in Hilversum. During the workshop, about 60 people attended from different supervisory agencies and with different roles (inspectors, managers, team leaders, and advisers). During the one-hour workshop, the researchers (the first two authors) assumed the role of facilitator. We started the workshop with a brief introduction to Generative AI. We introduced a definition of GenAI based on the *Overheidsbrede Visie Generative AI*. This definition reads: Generative AI is a form of AI that is capable of generating content such as text, audio, images, computer code, and videos (Ministerie van Binnenlandse Zaken en Overheidsrelaties, 2024) To provide context for this definition, we noted

that AI chatbots are the most recognizable applications of GenAI. We also explained that these can communicate through text in a way that closely resembles human interaction, with ChatGPT being a good example. After this introduction, participants discussed statements and questions in groups of three to five people (see Table 14), which we projected on the screen. We also asked the groups to complete a form with the statements and additional questions such as: Yes/no?; Why yes?/Why not?; How does it contribute/not contribute? These additional questions were intended to capture participants' deeper considerations and encourage them to think more nuancedly. After discussing a statement, the researchers facilitated a brief plenary discussion. This aimed to share the groups' insights so that attendees could learn from each other. After discussing all statements and the question, the researchers briefly presented three insights that had already emerged in the research up to that point

Table 14: Statements and question workshop 'Generative AI for supervisors'

Statement and questions discussed	
Statement 1	Generative AI can contribute to the quality of my work.
Statement 2	Generative AI can replace 'professional intuition'.
Statement 3	Generative AI can contribute to the creative process.
Statement 4	I need my own professionalism to make good use of Generative AI.
Question 1	What are the key conditions that contribute to successful use of Generative AI within inspections?

6.3.2. Data analysis

The data analysis included seven transcripts of the semi-structured interviews, including one double interview, and 15 workshop documents containing statements and related questions. The documents containing statements were manually processed into digital documents before analysis. For the analysis, ATLAS.ti, a specialised software for analysing qualitative data, was used. The data was analysed by coding. The first author performed most of the coding. During coding, he read the transcripts or workshop documents, while marking the most important

statements. He applied open coding, meaning the coding was done without preconceived themes, but followed the data. Open coding helps identify emerging themes within the data (Williams & Moser, 2019) . After a first round of coding, the authors jointly discussed the result of this first round of coding. They exchanged their interpretations and observations. Based on this exchange, the coding could be further refined. An example of codes that emerged included 'professionalism of user' and 'hiring and training people with AI knowledge'. We named the underlying theme 'Requirements for GenAI use' . We used a similar process for other codes, with themes such as 'Disadvantages of GenAI' and 'Advantages of GenAI' emerging. As the themes were relatively varied, we decided to present the main lessons, which, in our opinion, best represent the data. We present these lessons in section four.

6.4. Results

In this chapter, we describe key findings of the study, through describing eight main lessons. We illustrate these lessons through quotes from the interviews and workshops. The quotes were mentioned during the interviews, or written on the workshop forms. Quote can be identified by a [W] for workshop, or an [I7], when it concerns interviewee seven, for example. We try to stick as closely as possible to the empirical material within this section. In section five, we will reflect on the empirical results and link them to concepts from the literature.

1. Professionals see both substantive and process gains from GenAI
The professionals note that GenAI brings both substantive and process benefits. In terms of content, they mainly see gains in being able to think along and identify possible missed perspectives (see also point three). An inspector can, for example, ask a GenAI tool to critically examine their approach to an inspection and for suggestions on how this approach can be strengthened. This increases the quality of the content of their work. At

the same time, professionals see and hope for gains in time and efficiency - we call them process-oriented gains because we are talking about the work process. This involves, for instance, reporting or generating (parts of) reports. Through this efficiency gain within the work process, they hope to have time left over for the 'human' aspects of the work, such as human contact with colleagues or inspectees, depending on the role within the organisation.

- *'You create extra time for other things.'* [W]
- *'It also delivers more content-wise.'* [I2]
- *'If it's going to save time and you can spend that time on other things. That it's also going to have a profitable effect.'* [W]

2. GenAI can complement and potentially make professionalism explicit

Professionals have characterized GenAI in different ways, such as sparring partner, additional colleague, or mirror. What all these characterisations have in common, and what the professionals collectively appoint, is that GenAI will be mainly complementary to existing professionalism, but never a complete replacement of it. Some professionals also comment that GenAI could potentially be a tool that exploits professional intuition. By this they mean that GenAI could possibly help give words to the 'feeling' of supervisory professionals, or link their intuitions to certain concepts. In doing so, it could make this tacit knowledge more explicit. A mentioned example relates to conducting an inspection visit where there are multiple areas for improvement, compared to a visit where everything appears to be perfect. Something in the inspector's intuition suggests that something is not quite right about the second inspection visit, but the inspector cannot find the right words to express their concern. They hope that by interacting with GenAI, they are indeed be able to articulate this concern.

- *'It can help make feelings more concrete.'* [W]

- *'No replacement, but as support for your professionalism.' [*W
- *'More know more than one.'* [W]

3. Generative AI is creative, but only based on existing knowledge
Professionals within the study almost collectively note that GenAI can also contribute to work that requires creativity. They mainly mention that GenAI can help them find 'different angles' during their work. This can include, for example, angles that are unusual for professionals, or that they do not normally think of when doing certain tasks. This is also illustrated by the fact that 'out-of-the-box' thinking was mentioned a number of times. For instance, according to respondents, GenAI can be used to create an 'outside view' of a situation, which is a bit further away from the normal way of looking at the situation. One of the mentioned examples comes from youth care supervision. A common way to speak with clients of youth care institutions is to ask managers ultimately responsible at the institution. During a physical visit supervisory professionals ask these managers which children are open to a talk. However, this carries risks, as it means only those who have a good relationship with those ultimately responsible may dare to speak up. An interviewee suggested that a simple question to a GenAI system could have potentially alerted inspectors to this issue, thereby providing a new perspective on how to organize this differently in the future. At the same time, respondents also see that there are limits to the creativity of GenAI, which is, after all, based on existing knowledge. The 'brilliant mistakes' a generative AI tool is not likely to make - the brilliant mistake is the phenomenon where someone sometimes accidentally discovers something valuable or useful, without looking for it.

- *'It gives inspiration, different perspectives. '* [W]
- *'Helps with out-of-the-box thinking. Is actually an extra colleague.'* [W]
- *'It's all based on existing knowledge, so conservative.'* [W]

4. Professionals see AI as a catch-all term

Professionals within the study struggled to distinguish between different types of AI. Although the focus of this study is on GenAI, several times within the interviews they referred to AI, rather than GenAI. In addition, several interviewees indicated that they were unsure about the type of AI they were talking about at the time. Thus, the limited knowledge of AI means that AI degenerates into a catch-all term. As a result, respondents sometimes attribute features or capabilities to GenAI, for example, that these types of tools or currently do not have. For instance, multiple respondents mentioned that they hope GenAI can be deployed for a core task of supervisory agencies: determining where resources should be allocated based on risk assessment. The fact that professionals have difficulty distinguishing between different types of AI may point to methodological problems, since the findings would then no longer be specific to GenAI. Although these are legitimate concerns, respondents generally did appear to be aware of what GenAI can actually do. They simply estimated its capabilities more broadly than what is currently possible. This seems to justify attributing the remaining findings to GenAI.

- *'Maybe something like that could be handy. But I really don't have enough knowledge about it yet.'* [I8]
- *'But I don't know if it's generative anymore. I think that becomes almost decision-making AI.'* [I3]
- *'Yes, but whether that's AI, I don't know.'* [I4]

5. Professionalism is a prerequisite for good GenAI use

Professionals largely indicated that they expect professionalism to be a prerequisite for GenAI use. In other words, the knowledge, skill and experience of the professionals is needed for the GenAI tool to be truly valuable. The professionals see the importance of this professionalism in two places: 1) in questioning the GenAI tool (*prompting*), 2) in assessing the outcome of the GenAI tool. Professionalism is needed for querying because the outcome of GenAI tools is highly dependent on correct input.

Because of the experience a professional has, they know exactly the right questions to ask. Subsequently, this professional also has the experience and knowledge to assess the output. Precisely because of this experience, the inspector, for example, is able to see that the GenAI has hallucinated, or does not have the nuances quite right.

- *'My own professionalism remains necessary for the right direction.'* [W]
- *'You need basic professionalism to ask good questions.'* [W]
- *'To be able to give the right inputs and value the outputs.'* [W]

6. Not everything of interest can be captured as data for the GenAI tool
Several professionals indicate that not everything of interest can be captured as data on which the GenAI tool can be trained. Many signals on which, for example, an inspector, but also another professional, determines choices come from signals that are not directly explicit in nature. Here, for example, non-verbal communication is mentioned. It is precisely non-verbal communication that may contain bits of information or 'signals' that a GenAI tool cannot capture so far. Precisely the type of signals that emerge in social contact, for example, but lie just below the surface, feed what is known as intuition or professional intuition. This involves, for example, how a supervised entity reacts when inspectors come by for an inspection visit.

- *'We as humans can perceive non-verbal communication for example.'* [W]
- *AI cannot replace energetic contact. There are signals in there that feed your intuition. Things you can't interpret or pick up on and can't capture in AI.* [W]
- *Intuition is about much more than experiences than you can capture even in an LLM, for example. Not everything is (already) data.* [W]

7. Professionals want safe space to experiment

Several professionals indicate that they would like to experiment with generative AI in a safe environment. When the professionals talk about a safe environment, several things are mentioned in this context. In particular, they refer to the problem of data. Public GenAI tools like ChatGPT - more or less banned for civil servants - are seen as not safe for two reasons. First, because it is unclear what happens to data fed to the system. There is a fear that data fed to the system will end up in places where it should not. There is also ambiguity about the origin of the data that current public systems are trained on. Thus, the idea of security particularly refers to data. Security is also associated with the consequences of using GenAI. People are afraid to use it in 'real work' because the consequences of decisions are often large. So in this context, security seems to refer mainly to a closed environment used to experiment with internal data on how GenAI systems work. That experimentation in a secure environment is interesting for another reason (see point eight).

- *'So you also have to be able to go about your business now in a safe environment. So those are kind of important things in this.'* [I1]
- *'And that safety has to be guaranteed, of course. You can't start experimenting a bit where you then suddenly unleash a crisis, somewhere.'* [I1]
- *'I would love to experiment with that in a closed environment where it's guaranteed not to be out on the street.'* [I5]

8. It is precisely by using GenAI that you learn the risks , and don't fall behind inspectees

Several professionals indicate that it is important to use GenAI because only then do you learn about the risks. When it is not possible to use GenAI, you also do not learn what the tool can and cannot do, and how to properly use it yourself. It is also indicated that learning about the risks is very important for another reason, namely 'the outside world'. Moreover,

if inspections cannot use GenAI, and thus do not learn anything about the practical risks, opportunities, and ways to deal with GenAI, they also do not understand how the environment of the inspection can use GenAI. As an organisation overseeing other parties, it is important to understand what these parties can potentially do with a technology like GenAI. This is impossible if the use within central government is restricted. There could even be the risk of 'falling behind'. In the context of supervision, respondents note that falling behind can have significant implications, as it can compromise the quality of supervision when you don't understand how the outside world is dealing with GenAI.

- *'Because you find out the risks precisely by taking a little risk anyway.'* [I8]
- *'But at the very least, you need to know what is possible so that you can also appreciate the rest of what happens in the outside world.'* [I1]
- *'Well, you're not going to recognize the risks and so you immediately fall behind. And you just also put an innovation development at a standstill with that.'* [I8]

6.5. Discussion

In this section, we connect the empirical findings described in chapter four to our conceptualization of supervisory professionals. In this way, we reflect on the impact of GenAI on supervisory professionals and explore possible future scenarios.

6.5.1 The impact of GenAI on *tacit knowledge* and professional intuition

The first characteristic referring to public professionals is the strong reliance on tacit knowledge or professional intuition (de Bruijn, 2012). The empirical results provide, on the one hand, a picture that GenAI can help

make tacit knowledge explicit. GenAI may have the language or words to do so. GenAI tools can serve as a sparring partner, helping to make professionals' knowledge explicit. After all, the GenAI tool is linguistic, and can find the 'right words' based on the right prompt.

On the other hand, the role of tacit knowledge remains. The main reason is simply that tacit knowledge does not always lend itself to explication. Indeed, if supervisory professionals start relying heavily on GenAI tools, this might weaken their professional intuition. When the professional is heavily guided by a tool, they may come into less direct contact with complex situations in which precisely their intuition and professionalism are nurtured and developed.

In addition, tacit knowledge is needed to write the right prompts for the GenAI tool. After all, a layperson is less able to write the right *prompts* that fit the language and way of doing things used within the supervision domain or within the role the supervision professional holds. That same implicit knowledge is needed to assess the output. Indeed, knowledge of the context is needed to determine whether what the GenAI generates is correct, or not. The second observed reason why supervisory professionals see a role for tacit knowledge is because the professionals are able to observe so-called *telltales*, while a GenAI system cannot. In Dutch, a *telltale* refers to a sign that reveals something that would otherwise remain hidden. Professionals indicate that they sometimes encounter such signs in their work. This could be, for example, the behaviour of a supervised person.

6.5.2 The impact of GenAI on autonomy and discretion

The second characteristic of supervisory professionals is the strong degree of autonomy and discretion to make decisions about and in their daily work (Lipsky, 1980) . Again, supervisory professionals see a role as a sparring partner for GenAI tools. That supervisory professionals are autonomous and act within their discretionary space also means that they

sometimes have to be creative. They have to be creative in dealing with the complex situations they face. GenAI, according to our respondents, in these situations can serve as an extra colleague that encourages you to think out-of-the-box.

At the same time, there are also risks with GenAI for autonomy, discretion, and creativity. First, in fact, our respondents see GenAI as 'conservatively creative'. We mean here that people mostly mention that GenAI is creative, but to a certain extent. The professionals mention that GenAI is trained on existing knowledge, and thus is never really innovative. When using GenAI, the professionals see a small chance of serendipity - an opportunity for the brilliant mistake. GenAI is designed to give the best possible answer based on a *prompt* - and not to find something innovative that the user was not looking for (the definition of serendipity). These insights about creativity also led to larger and more philosophical questions about creativity that are in tension with the conclusion in section 5.1. Isn't human creativity to discover new things also based on the knowledge that already existed, but combined in new ways? In other words, isn't GenAI 'just' as creative as the professional?

The second risk is that the use of GenAI curtails the autonomy and discretion of the supervisory professional. Theoretically, the GenAI can impose guidelines on how to act. With high confidence in the technology, this may lead professionals to rely less on their 'own discretionary space' in their actions, but to have it partially filled by the GenAI. The possible influence on autonomy and discretion also shows in thoughts about *prompts*. After all, if good *prompts* are prescribed, where is the freedom of professionals to deviate from those guidelines? This tension between GenAI and autonomy can be further reinforced by specifying 'from above' what GenAI may or may not be used for. At the same time, reality does not seem to be moving that quickly. Given the importance of professional discretion for both creative prompts and evaluating output, the path toward a system-level bureaucracy, as discussed by Bovens

and Zouridis (2002), does not yet appear to be at hand in times of GenAI. Furthermore, our empirical work as well as the work of Selten, Robeer and Grimmelikhuijsen (2023), who show that police officers only follow AI recommendations if they confirm their professional judgment, provides reason to argue that even with GenAI, supervisory professionals will mostly first rely on their professional judgment

6.5.3 The impact of GenAI on dealing with conflicting perspectives

The third feature from the conceptualization is the fact that public professionals within supervisors have to deal with conflicting perspectives and values. Again, our respondents see a role for GenAI. Some respondents mentioned that GenAI can sometimes serve as a sparring partner who can name perspectives not yet viewed. We can apply this logic to the conflicting perspectives that professionals have to deal with. In a sense, some see GenAI as a possible representative of perspectives not yet considered. GenAI can, as it were, serve as a 'contradiction' or *countervailing* power and counterbalance the professional's view. After all, this view can sometimes be so internalised that the professional no longer notices his own view of reality. In this way, GenAI can ensure that decisions made by supervisory professionals do more justice to different perspectives.

Even when dealing with conflicting perspectives, however, there are risks in relying too heavily on the GenAI tool. First of all, the GenAI tool cannot actually weigh up perspectives. The complex reality of supervisory professionals requires sensitivity to the context in order to determine which perspectives matter and how to weigh them. The latter in particular is something GenAI cannot do.

The difficulty of dealing with conflicting perspectives can also be seen in our empirical results on whether GenAI can be used by supervisors. Professionals show strong awareness about the importance of 'safety'. For example, they mention the need for safe experimentation space. It

seems that when it comes to GenAI, professionals are strongly driven by the values of 'security' and 'privacy'. A possible explanation here lies in the fact that there have been many warnings about the commercial nature of public GenAI applications such as ChatGPT. This seems to have made professionals aware that it is unclear where the data you put into such applications goes. At the same time, there is also a realisation that professionals or the regulator as a whole must remain innovative and progressive. This is where an inherent tension between perspectives reveals itself, as additional security may affect the extent to which innovation and experimentation can take place.

One perspective which is conspicuously absent within our empirical results is that of legitimacy. Especially within the context of supervisors, where decisions with relatively large consequences often have to be made, the impact of GenAI may influence the legitimacy the supervisor has. This legitimacy also arises in the relationship between professionals (mainly inspectors) and supervised persons.

6.5.4 The impact of GenAI on the relationship between professionals and inspectees

The fourth characteristic of professionals based on the earlier conceptualization is that supervisory professionals carry out their work in relation to inspectees. Here again, the role of GenAI is interesting. A number of respondents indicated that GenAI might serve as a complement, as well as provide substantive gains. This logic may also hold true when looking at professionals and the strategic behaviour others may show. GenAI can help supervisory professionals gain much more detailed insight into the behaviour of others around them. At the same time, there are also things that GenAI cannot do. Especially when it comes to strategic behaviour, GenAI is unlikely to be able to see what game is being played. For example, the 'game' between inspector and inspectee is highly context-dependent. Insight into a specific contact

GenAI cannot provide. So with that, again, GenAI has a role, but it also certainly has limitations.

It is also important here that inspectees have every opportunity to use Gen AI. Almost all supervisory professionals indicate the importance of using GenAI to learn about risks and opportunities. There is a concern that inspectors and other supervisory professionals do not have as good an understanding of what the technology can do as those organizations they are tasked with supervising. The logic adhered to by professionals within the study is that it is more important to take risk, and understand what the technology is capable of, than to take the risk of falling behind when using these kinds of technologies. Indeed, falling behind can affect the relationship, and thus put pressure on the quality of monitoring. Failing to understand what GenAI can do can thus touch on the legal task of supervisors and how they perform it qualitatively.

As mentioned earlier, there is another risk with the use of GenAI and a possible effect on the relationship between professionals in inspectorates. It is notable that this was barely mentioned within the study. We are referring to the legitimacy of supervisors. It is legitimate to ask what the effect is on inspectees when supervisors partially rely on GenAI. Will this affect how legitimate these inspectees see the decisions of the supervisor?

6.6. Conclusion

The main question of our research is: *How do professionals within supervisors expect generative AI to affect their work*. Our research on four characteristics of professionals (reliance on tacit, '*tacit*' knowledge, their high degree of autonomy, their regular confrontation with competing values and their relationship with inspectees) shows that supervisory professionals expect their work to be affected by the emergence of GenAI.

GenAI is gradually changing from a reference tool to a sparring partner, and thus the relationship between supervisory professionals and GenAI is changing from a one-sided to a multi-sided relationship. When GenAI is used as a reference tool the relationship is on-sided. The professional asks a question, GenAI provides an answer. However, we also see a multi-sided relationship emerging between professional and GenAI. [The professional provides input and receives not only an answer, but also suggestions or alternative perspectives. This requires a critical view from the professional based on existing professionalism. This creates a process where the GenAI tool and the professional challenge and potentially strengthen each other. This development of GenAI as a sparring partner has at least three implications.

- It broadens the possibilities for the professionalism of supervision. For example, it can lead to making *tacit knowledge* more explicit, to more creativity, to a better understanding of behavioural patterns of inspectees.
- It places high demands on the professionalism of supervision. For GenAI to be used properly as a sparring partner, it also requires highly developed professionalism of supervisory professionals. Using the right *prompts* and valuing the output of GenAI requires the knowledge and skills of one professional. The more powerful Gen AI becomes, the stronger that knowledge and skill must be developed.
- This development comes with risks to professionalism. As with human sparring partners who think along and against, there is the risk of the tool who thinks along and against that becomes too dominant compared to the professional. Using GenAI can also lead to loss of tact knowledge. It can limit creativity. The weighting of perspectives and the interpretation of the relationship with inspectees is almost always context-specific - and the use of GenAI is thus highly risky.

GenAI as a sparring partner, with its requirements and risks, implies that dealing with GenAI is a learning process. There is a need to learn not only about the practical use of these tools, but also about the above three points. Under what conditions are the opportunities for professional supervision broadened? What requirements does the use of GenAI place on the professionalism of supervision? What are the risks, what are no go areas for using GenAI?

The above implications also affect various professional roles of supervisory when using GenAI.

- *The supervisor as public professional:* Public professionals are expected to embody and represent important public values such as transparency, legitimacy, and privacy. With GenAI as a sparring partner, safeguarding these becomes even more important due to its greater role. The question is which values GenAI represents and whether these align with what is expected of supervisory professionals.
- *The supervisor as risk-sensitive professional:* Risk-based selection is inherent to the work of supervisory agencies. Risks can be detected based on data and observation. GenAI as a sparring partner can strengthen risk sensitivity by challenging the inspector to use the right prompts and maintain healthy scepticism toward generated results. GenAI as a reference tool can weaken risk sensitivity, especially if GenAI is used as an alternative to observation.
- *The supervisor as responsive professional:* Here, responsiveness refers to the operational relationship with supervised entities and other parties. Van Erp and van Wingerde identify this role, although they primarily take an administrative perspective (2013). If supervised entities can use GenAI more effectively (as a sparring partner) than the supervisor, then the supervisor's position may be weakened.

For regulators, it is therefore important to find ways to deal with GenAI. There is a key role for learning. We make three recommendations about this learning process. First, explicitly start with making space to experiment with internal GenAI tools. In this way, professionals can learn about the role GenAI could have in the work process. This may also help professionals learn to better recognize output generated by GenAI. But how and by whom that learning process can be shaped? The second recommendation is to train supervisory professionals who have both GenAI knowledge and domain-specific knowledge. These professionals must have the tacit knowledge to be able to use GenAI valuably. Finally, we advise policymakers and managers to develop clear guidelines for responsible GenAI use within supervisory agencies. Let professionals experiment, but make clear where GenAI may and may not be applied. This way, the learning process can focus on those aspects of professionalism where value can actually be created.

GenAI that is increasingly being deployed as a sparring partner is a delicate process. GenAI can strengthen professional work when it serves as a sparring partner. In this process, however, it is essential to ensure that professional tacit knowledge is preserved. Here lies a challenge for supervisory organizations to find a balance between technological innovation and core professional values

6

Discussion is an exchange of knowledge;
argument an exchange of ignorance.
– Robert Quillen

7

Discussion and Conclusion

The previous three chapters each answered one of three sub-questions through different papers. This chapter connects these findings to answer to the main overarching research question. In doing so, I maintain the core insights from each paper. However, the analysis and answer provided here provide additional insights that emerge through comparing and integrating findings of the papers.

7.1. Answering the main question

Here I consolidate the findings of the papers, and provide an answer to the main research question that this dissertation posed in the introduction. The question start from the premises that AI algorithms transform organizational processes while organizational practices simultaneously shape AI adoption and use. The main RQ is:

> *What are the organizational implications of the adoption of learning algorithms in public organizations?*

The short answer is that AI adoption has major and mixed implications for public organizations. When adopting AI public organizations deal with increased organizational complexity because of the increase in types of professionals and number of interactions between professionals. The increase in variety and complexity creates misalignment between current organizational structures and those required for effective AI adoption. This misalignment is illustrated by several phenomena. These phenomena highlight how public organizations start to transform with AI adoption. Key phenomena are:

- The emergence of novel boundaries (chapter 4). Boundaries emerge as a consequence of facilitated interaction between developers and end-users. While that facilitated interaction is focused on blurring boundaries, the opposite also occurs.

- Discretionary dependencies between various professionals (chapter 4 and 5). AI adoption leads to a shift in discretionary power from end-users, i.e. incumbent professionals, to other organizational actors. The shifts happens because of the various decision-making roles that exists and emerge within AI development (van der Voort et al., 2019; Zweig et al., 2018).
- High time investments without clear returns (chapter 4 and 5). Interaction to promote AI adoption does not happen by itself. It requires time and effort, sometimes leaving less time for core activities, for example activities done by incumbent professionals. However, the fruits of interaction cannot always be foreseen in advance, both for managers and professionals themselves.
- The need to acquire new skills (chapter 4, 5, and 6). AI adoption requires new skills and novel ways of thinking. Exemplary is the need for end-users, e.g. incumbent professionals, to at least have a basic understanding of AI. However, for successful AI adoption, also the skillset of developers seems to broaden beyond only technical expertise.

The occurrence of these phenomena support the message that organizations underestimate the impact AI has. As noted before AI is not a stand-alone technical tool (Lorenz, 2024; Meijer et al., 2021; van der Voort et al., 2019). AI is intertwined with the organization, professionals, their interaction, and more. AI has major effects that public organizations are not immediately aware of.

However, not only unforeseen phenomena that illustrate misalignment occur. Other – potentially temporary – phenomena occur that signal that organizations are already becoming more capable in dealing with AI.

- Learning processes occur (Chapter 4, 5, and 6) . Adopting AI demands the development of novel skills. Both incumbent

and novel professionals expand their skillset with AI adoption. Incumbent professionals start to learn about AI on a technological, social, and organizational level. They are beginning to understand the technology, the novel roles accompanying AI adoption, and the potential implications better and better. Similarly, developers expand their skillset beyond technical expertise. Looking at the entire organization, they are becoming more mature in adopting and using AI.

- New roles emerge bridging different - currently separate - professional groups (chapter 4 and 5). AI adoption contributes to the emergence of boundary spanners, individuals who are engaged in boundary spanning activities, processes, and tasks (Williams, 2013). They translate concepts and ideas between professional groups, most notably developers and end-users, thereby having a critical role within AI adoption. In today's organizations, official positions do not exist yet. Therefore, these boundary spanners are most often organizational actors who grow into this role, because being a boundary spanners is not a profession in literal terms.

With these key phenomena, short-term misalignments, possibly seem to be a necessary evil to move towards an AI-driven organization. An organization wherein AI takes its place alongside incumbent and novel professionals, who know more and more precisely where AI does and does not add value. An organization wherein professionals provide counterbalance to each other and AI systems where necessary. With this answer, we turn our attention to the sub-questions of the separate papers. After that, section 7.3. will highlight more detailed lessons arising from answering the main research question.

7.2. Revisiting the sub-questions

7.2.1. Research question 1: Boundaries between developers and end-users

Research question one asks: *How does co-creation influence boundaries between developers and end-users in ML development and how do boundaries behave as a result?* We tried to answer sub-question 1 through a case study with ethnographic elements. By means of observation, document analysis, semi-structured interviews, and presence at events and meetings we studied intentional interaction – facilitated by managers - between developers and end-users and how that impacted boundaries between them and within the organization. We drew from organization science theory (e.g. Bechky, 2003; Carlile, 2004) that discusses knowledge sharing between occupational communities and difficulties that might arise. We used that theory to understand knowledge sharing difficulties between ML developers and end-users.

Chapter 4 suggests that co-creation has mixed and unpredictable effects on boundaries between ML developers and end-users. To use the words of this comprehensive essay, organized interaction between developers and end-users within AI development, has various effects.

- Organized interaction indeed leads to blurring boundaries. Developers and end-users involved start to speak a shared language. They work towards the same goals, that they have jointly negotiated and agreed on.
- However, differences persist on a deeper level. Although developers and end-users work towards the same goal, it appears that the shared language mostly relates to the organized interaction and is on a more 'generic' level. Within the context of the AI project, the actors understand each other. For example, both want to work on an 'AI tool', which

will help the end-user. However, fundamental differences persists. For example, in their understanding of concepts or the role AI will eventually play. For example, even after working together for a long time, it turned out that both actors meant something different by a 'tool'.

- Also, novel boundaries emerge between different end-user types because of organized interaction. Boundaries emerge between the end-users involved in the organized interaction, and those that were not involved. They too now speak a slightly different language, and their goals may differ.
- Hence, when organizations decide to facilitate interaction between diverse professionals, they may unintentionally generate 'waterbed effects', where blurring boundaries in one place, triggers the emergence of boundaries in another place.

Given the described effects of organized interaction, optimism around the necessity of interaction for AI adoption could be tempered. Interaction indeed blurs some boundaries between developers and end-users, but it simultaneously creates new ones and leaves fundamental differences to persist. Therefore, organizations looking to blur boundaries for the purpose of AI adoption, must consider both the intended benefits as the unintended consequences.

7.2.2. Research question 2: XAI as a learning process

Research question two asks: *What are the benefits and challenges of an interactive and iterative development approach for explainability of AI models and how do benefits and challenges occur?* We answered the question through action research, leveraging data collection through presence at meetings and a workshop, semi-structured interviews, online participant observation, and taking field notes. We tried to understand the effects of interacting throughout the development process between various professionals on explainability of AI models. We were interested in the these effects, since explainability is often seen as a desirable trait

of AI models. The literature has discussed that interaction may contribute to increased AI explainability (Bhatt et al., 2020; de Bruijn et al., 2021; Longo et al., 2024).

Chapter 5 highlights that an interactive and iterative development approach has several challenges:

- *Conflicting explainability demands.* Actors different preferences for the type and form of explanation can hinder understanding, especially during joint interactions.
- *A lack of the appropriate skills among developers.* Developers lack experience with the skills needed for interactive AI development, as their traditional approach prioritizes data acquisition, exploration, and experimentation.
- *The involvement of required stakeholders.* Securing the involvement of required stakeholders proved difficult because of unwillingness to be involved, or a lack of understanding for their necessary involvement.
- *Magnitude of time-investment:* The time-intensives of the approach was major, while the results were unclear at the beginning of the project.

However, interactive and iterative development also has benefits and contributes to explainability in multiple ways:

- *Personalization of explanations.* Through the possibility of differentiation between stakeholder involvement levels the developers could cater to varying needs and wishes.
- *Learning processes among developers.* Continuous interaction and iterations helped the developer team to learn novel skills and about the organization and its functioning

- *Selective engagement of critical stakeholders.* The flexibility ingrained in the development process gave the opportunity for the emergent presence of a critical stakeholder.
- *Organizational learning processes.* The constant iteration and interaction between developers and the organization stimulated and compelled them to start learning about and with AI.

While intensive interaction between external developers and organizational professionals has many challenges and does not directly affect the explainability of the AI model itself, it does contribute to a better understanding of AI within the organization and sets learning processes in motion regarding what is needed to adopt AI.

7.2.3. Research question 3: Generative AI and professionalism

Research questions three asks: *How do professionals within inspectorates expect generative AI to affect their work?* We tried to answer the question through semi-structured interviews and a workshop. We studied the potential effects of GenAI on the work of professionals working within regulatory agencies. By drawing from theories from several closely related disciplines, we conceptualized professionals working within regulatory agencies according to their autonomy (de Bruijn, 2012) and discretion (Lipsky, 1980), reliance on tacit knowledge (Howells, 1996), dealing with conflicting perspectives, and relationship with inspectees (De Bruijn & Ten Heuvelhof, 2019).

Chapter 6 suggests that professionals have mixed expectations about the effect of GenAI. Some expect a major impact, while others emphasize that some aspects of professional work can never be replaced by GenAI. Our reflections in Chapter 6 on the professionals perspectives highlight that there seems to be both a single-sided and a multi-sided relationship emerging.

- The single-sided relationship refers to the use of GenAI as a kind of reference tool. When GenAI is used as such, it is not very different from a search engine. The relationship is one-sided: the professional asks a question, GenAI provides an answer.
- The multisided relationship refers to the use of GenAI as a sparring partner. In this view, professionalism is seen as a prerequisite for good GenAI use. Professionalism and tacit knowledge are essential, because they are needed to know how to get the most out of GenAI. However, GenAI also pushes back on professionalism and tacit knowledge. GenAI may be a countervailing power to the assumptions underlying professionals' tacit knowledge.

While GenAI is seen described as a tool that may contribute to repetitive tasks, the emerging multi-sided relationship between GenAI and professionals highlights that GenAI may also impact complex tasks and the reliance on tacit knowledge that professionals often have.

7.3. Lessons related to the lenses

This section puts attention to more specific lessons related to the three lenses presented in the introduction. Table 15 gives an overview of the lenses and main lessons. These lessons are a further elaboration of the answer to the main question.

Table 15: Lenses and main lessons

Lenses	Lessons
Lessons about AI and professionals	1. Increasing variety in professional landscape 2. The emergence of the professional boundary spanner 3. Discretionary dependencies and re-distributed discretion 4. Inconclusive importance of tacit knowledge
Lessons about AI and interaction between multiple organizational actors	1. The double edged sword of interaction 2. The emergence of novel boundaries 3. The interaction skillset as a problem 4. The undesignable nature of interaction
Lessons about AI and the organizational context	1. Contextual tensions and the need to keep up 2. The need for contextual expertise

7.3.1. Lessons about AI and professionals

At the start of this dissertation I introduced the professional lens. Furthermore, the potential implications of that lens were discussed, including the AI-tacit knowledge force field, the emergence of new professionals, and shifting discretion and autonomy. Here, I discuss the lessons learned about AI and professionals, given what has been discussed throughout the main chapters of this dissertation. I will highlight four key lessons that have emerged throughout:

1. Increasing variety in professional landscape
2. The emergence of the professional boundary spanner
3. Discretionary dependencies and re-distributed discretion
4. Inconclusive importance of tacit knowledge

1. Increasing variety in professional landscape

As theorized, AI adoption necessitates the introduction of new types of professionals, such as data scientists or privacy experts (Lorenz et al., 2022; van der Voort et al., 2019). What has emerged throughout this thesis is that AI adoption over time dramatically increases variety in the

professionals landscape. AI adoption leads to three distinct developments: a) increasing variety in types of professional roles increases; b) increasing variety in necessary skills increases; c) increasing variety within professional roles increases. The fact that this development take place over time also makes it difficult for organizations to deal with them.

a. Increasing variety in types of professional roles

The empirical work throughout this thesis reveals that variety in the types of professional increases with AI adoption. For example, for AI to be of any use throughout its development a data scientists has to be involved. However, because of the fact that AI learns from often privacy-critical data, a lawyer or privacy specialist also is involved. But, the AI systems also need to be hosted. Hence, someone with experience in the field of ICT architectures is also involved. Even more examples could be given. When viewed from a slightly higher level, three categories of professionals seem to emerge: the end-users, the facilitators, and the mediators. As a consequence of this trend of growing variety in professional roles the number of 'blocks' within the organization and the chain of AI development and use grows. While the promise of AI is efficiency and reducing complexity, organizations grow increasingly complex.

b. Increasing variety in necessary skills

The increasing variety in number of professional roles may give the impression that this leads to job specialization, which involves breaking down tasks in to smaller simpler components leading to narrowly focused work (Mintzberg, 1979). However, it seems that while variety in types of professionals grows, also the number of knowledge some of these professions need is growing and the broadness of tasks that these professions have to do. Hence, in practice, *job enlargement* takes place (Mintzberg, 1979).

This point is illustrated trough the enlarged role of data scientists throughout the dissertation. The necessary skills the data scientists

needed to possess enlarged quite substantially. Besides skills relating to technical aspects of the AI systems, data scientists also needed questioning skills, project management skills, and skills that help them to involve and engage organizational actors.

c. Increasing variety within professional roles
While job enlargement takes place on the professional role level, simultaneously, specialisation takes place within professional roles. AI adoption leads to specialization within the earlier mentioned blocks (e.g. the different job roles within the organization dealing with AI). It seems to occur because of differences in expectations, potentially leading to various distinct groups within professional groups.

For example, some end-users actively engage with AI experimentation and use, while others hesitate to get involved due to scepticism about AI's value or a lack of knowledge. Similarly, some data scientists are enthusiastic about user interaction, while others prefer to be mostly involved with 'technical' tasks.

2. The emergence of the professional boundary spanner
AI adoption in public organizations seems to contribute to the emergence of professional boundary spanners. Boundary spanners are referred to as individuals who are engaged in boundary spanning activities, processes, and tasks (Williams, 2013). These include translating concepts between professional groups, or being part of a multi-disciplinary team.

By *professional* boundary spanners, I mean that these actors have a critical role within AI adoption, and have specific characteristics that come in handy to connect various professional groups that have to interact. The fact that their role can almost be seen as a separate 'profession', with specific knowledge and skills, makes them *professional* boundary spanners. Simultaneously, boundary spanners are not a profession in literal terms, because their role is informal. Actors grow into this role,

but its informal nature means it is not part of the formal accountability structure within the organization.

The findings indicate that at least three types of boundary spanners emerged.

- First, internal boundary spanners working as developer connected developers and end-users within organizations. Such type of boundary spanner seems to resemble what Williams (2013) calls '*frontline professionals as boundary spanners*' with the difference that the boundary spanners we found are internally oriented, instead of externally.
- Second, transversal boundary spanners connected externally positioned developers with domain experts in the organization these developers worked for. These boundary spanners more closely resemble Williams (2013) frontline professionals, as they connect the organization to the outside world.
- Third, boundary spanners working as end-users emerged that connected end-users with developers. They resemble the first boundary spanners in this list, with the difference that they are end-users instead of developers.

What these boundary spanners had in common was that they understood both AI to some extent and the domain in which it would be applied. Looking at these examples, it seems that these boundary spanners naturally emerge as a consequence of the job enlargement that takes place. Because of job enlargement in types of tasks, naturally professionals occur that span boundaries.

Overall, it turns out that boundary spanners - people who speak the 'language' of multiple professional groups, e.g. the developer and the domain professional, are essential in public AI adoption. They are able to translate concepts between different professional groups, and can

understand and empathise with both worlds. At the same time, their emergence raises questions about the focal position such actors can take. Previous research has highlighted that officially designated boundary spanners attracted significant power, thereby impacting AI adoption and outcomes (Waardenburg et al., 2022). Similar questions arise based on this dissertation. Are boundary spanners not too important in AI adoption processes?

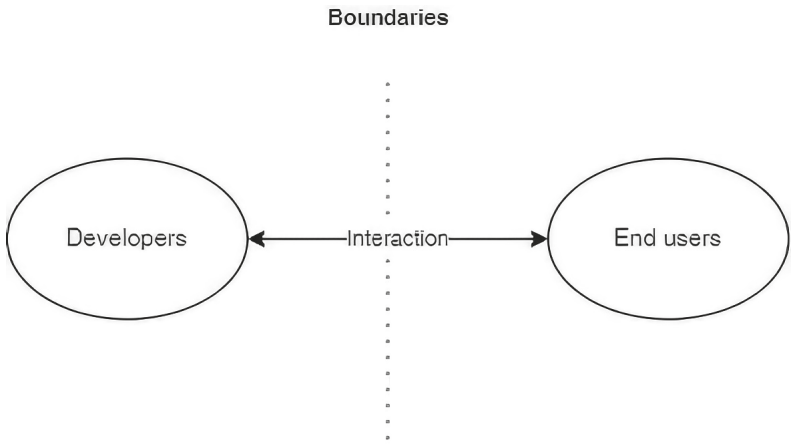


Figure 8: Illustration of boundaries between developers and end-users

3. Discretionary dependencies and re-distributed discretion

This dissertation reveals that AI adoption leads to a re-distribution of discretionary power from end-users to other organizational actors, as theorized in the introduction. Decision space moves from some actors , mostly incumbent professionals, to others. Paradoxically, the other organizational actors are dependent on the end-users domain knowledge to exercise their discretion, because end-users domain knowledge is often a key element needed for the development of helpful AI models. I call this dependency to exercise their discretion 'discretionary dependency'. This finding broadens our understanding of what discretion means in AI adoption, because existing work is more focused on discretion of the end-user or a discretion shift towards the AI model (Lorenz et al., 2022; Young et al., 2019).

The shifts takes place because of the various decision-making roles that exists and emerge within AI development (van der Voort et al., 2019; Zweig et al., 2018). For instance, privacy officers determine data boundaries based on legal frameworks, while data scientists select specific algorithmic approaches.

The fact that discretion is redistributed across other or novel actors has several consequences. First, end-users might resist the redistribution of discretion because they fear their role is curtailed or changed in unwanted ways (Delfos et al., 2022). Hence, the needed involvement of end-users, is complicated, as they might not want to contribute. Second, a disconnect may emerge between formal and real responsibilities. The organizational accountability structures may still reflect old decision-making patterns, while discretion has shifted to others that should actually be hold accountable.

4. Inconclusive importance of tacit knowledge

Some argue that professionals tacit knowledge becomes less important as AI is able to partially replace or embody it (Brynjolfsson et al., 2023). AI models, especially those based on data, would be able to see nuances and connections that were previously reserved for experienced professionals with extensive domain knowledge. This would significantly change the role of end-users, e.g. mostly inspectors.

However, the findings reveal that at least the professionals themselves expect the importance of tacit knowledge to be preserved. This is highlighted trough the following findings:

- Whether GenAI or ML, professionals indicate that there are always things that factor into their decisions that cannot (yet) be captured in AI models, giving importance to tacit knowledge.

- Also, domain experts were tremendously important in giving meaning to the data. Domain experts or at least people with intuitive knowledge about the domain in which the AI application would be implemented, on multiple occasions proved essential. One example is bringing the right segregation in the data for the AI model. Without the right segregation, the resulting AI application could have been useless.

In sum, while some indications exists that point into the direction of a decreasing important of tacit knowledge (Brynjolfsson et al., 2023), this dissertations empirical material teaches us that this might not be the case. On the contrary, there are good reasons to believe professionals' tacit knowledge is as important as ever with the growing adoption of AI.

7.3.2. Lessons about AI and interaction between multiple organizational actors

At the start of this dissertation I introduced the professional lens. Furthermore, the potential implications of that lens were discussed, including coordination issues, knowledge boundaries, and demands and uncertain outcomes. Here, I discuss the lessons learned about AI and interaction between multiple organizational actors, given what has been discussed throughout the main chapters of this dissertation. I will highlight four key lessons that have emerged throughout:

1. The double edged sword of interaction
2. The emergence of novel boundaries
3. The interaction skillset as a problem
4. The undesignable nature of interaction

1. The double edged sword of interaction

Some argue that interaction is necessary to develop 'helpful' AI systems (Lorenz et al., 2022; Mayer et al., 2023; Waardenburg & Huysman, 2022).

The thinking is that interaction ensures, among other things, that AI systems are to the liking of the end-users, but also that choices in the development process may be better substantiated. Interaction for more successful AI adoption thus seems to be a no-brainer.

However, this dissertation shows that the effects of interaction are not sufficiently recognized. Interaction indeed has positive effects, such as reduced boundaries between people interacting, for example developers and end-users. But at the same time, there are also undesirable or unexpected effects like the emergence of novel boundaries. With this, interaction for successful AI adoption is a double edged sword. It has positive sides, but most definitely also undesirable or unexpected effects. Two unexpected effects that this dissertation highlights below are the emergence of novel boundaries, and the interaction skillset as a problem.

2. The emergence of novel boundaries

Within AI development professionals from different occupational communities may struggle to understand each other when interacting, most notable developers and end-users (Waardenburg et al., 2022). Whereas the existence of boundaries is relatively unsurprising, the emergence of novel boundaries is. This dissertation argues that facilitating interaction to contribute to blurring boundaries, contributes to the emergence of novel boundaries.

Chapter 4 illustrated this phenomenon vividly. It shows that novel boundaries emerged as a consequence of facilitated interaction between developers and end-users. The ML system co-creation process brought together selected end-users and AI developers. However, formal requirements and limited project capacity prevented broader end-user participation, either by choice or circumstance. This selective involvement created an unforeseen novel boundary: over time, the gap between participating and non-participating end-users increased, creating a new boundary within the end-user group. Worth mentioning

is that the emergence of boundaries contributes to the emergence of boundary spanners (see 7.2.1. Lessons about AI and professionals). As new boundaries emerge, they create opportunities for boundary spanners who bridge them.

3. The interaction skillset as a problem
As mentioned above, interaction is cited as a necessity for successful AI adoption (Lorenz et al., 2022; Mayer et al., 2023; Waardenburg & Huysman, 2022). One of the unexpected effects is that interaction itself may be part of the problem. The interaction skillsets of the actors involved are underdeveloped or naturally different.

Chapter 5 shows that the developers involved needed an additional 'social' skill set to be successful in interacting with potential end-users. Having to learn an additional skillset contributes to a major extension of their professional role. The question becomes whether this can be reasonably expected form a single actor in the AI multi-actor network. In addition, interaction naturally takes place differently between different actors. Chapter 4 empirically shows that while interaction appears to be successful, different types of actors continue to reason differently or interpret words that appear unambiguous differently. Both issues raise the question of whether the right conditions are currently available to make interaction a resounding success.

4. The undesignable nature of interaction
The above described lessons highlight another lesson. Given that interaction is a double edged sword, as it leads to unexpected or undesired effects, the question rises whether the effects are controllable. Is it possible to design the interaction process in such a way that these effects do not occur?

This dissertation highlights that the interaction process between actors needed for AI adoption is more or less undesignable. This is because

there is an inherent unpredictability about the effects of interaction. It is difficult to estimate in advance what will happen, for example illustrated by the above lessons about emergent boundaries and the need for a novel skillset.

However, also chapter 5 provides additional illustrations. First, the time investment became unexpectedly rather large, while efficiency gains were unclear. Is taking additional time to co-create AI systems worth it, if the efficiency gains are only marginal? Also, the messiness of co-creating AI systems seemed unaccounted for. The interaction process highlighted that actors had sometimes contradictory wishes. Upfront, these effects may possibly be anticipated. But, then there is still the question of whether they can be prevented by designing the interaction process in a particular way. The answer to this questions, is probably no.

7.3.3. Lessons about AI and the organizational context

In the introduction of this dissertation I introduced the contingency lens. Furthermore, the potential implications of that lens were discussed, including the need for the context, the requirements of the context. Here, I discuss the lessons learned about AI and the organizational context, given what has been discussed throughout the main chapters of this dissertation. I will highlight two lessons that have emerged throughout:

1. Contextual tensions and the need to keep up
2. The need for contextual expertise

1. Contextual tensions and the need to keep up

As discussed, public organizations try to meet the conflicting demands from their wicked context as best they can. However, it is impossible to satisfy all these demands when adopting AI. Some contextual demands are inherently conflicting, for example explainability and performance of AI models (Burrell, 2016), while others may be jointly realized, but there is no agreement on the importance of these values. This dissertation

shows that in practice two findings can be observed about the demands of the context.

- First, *the asymmetry of AI adoption.* Due to some reluctance in AI adoption by public organizations, possibly due to many external demands, there is asymmetry between public organizations and the outside world. However, because of the rate of AI adoption at contextual organizations, public organizations are forced to go along. Indeed, when they don't do this, they run the risk of falling so far behind, that they cannot catch up with the environment. Not being able to keep up with the environment may mean that public organization are unable to properly perform their public duties and be a counterforce against the environment it is supposed to be.
- Second, the contextual values seem relatively implicitly incorporated in the adoption process of AI systems, while the environment may require the explication of values. The empirical findings across all chapters provide little reason to argue that there are very explicit value trade-offs. This finding may be explained by the lack of time that public organizations, given the risk of falling behind. However, given the asymmetry in AI adoption, some values are indeed more present than others. For example, an efficient and innovative government that stays up to speed.

Both findings highlight that the context is demanding. Public organizations seem moderately aware of AI adoptions' value conflicts. Because of the speed of contextual developments, this leads to a tension. On the one hand, there is the need to do AI adoption in 'the right' way, at the same time it may be impossible because of the need not to fall too far behind with the risk of being unable to perform their duties in the future.

2. The need for contextual expertise

Chapter 2 theorized that public organizations at times rely on external organizations for successful AI adoption, because they lack AI expertise themselves (Delfos et al., 2022). These external organizations have the knowledge on how to develop and use AI, and can help public organizations with this challenging task.

Indeed, public organizations may rely heavily on external parties for AI knowledge. With the help of external knowledge, these organizations can embark on learning processes that ensure the organization moves forward in its dealings with AI. Hence, the context can be seen as a source of help. However, this dissertation adds a tension to this finding.

The reliance on external parties may contribute to fragmented agency. Because of the reliance on external parties, responsibility for the quality and outcomes of AI systems are shared between external and internal actors. This creates tension, is the need for contextual expertise that large, that it justifies partial outsourcing of responsibility around key choices of public AI systems? And does this not make public organizations too dependent on external actors?

7.4. Practical recommendations: from boundary spanners to internalization

Section 7.3 presented specific lessons. Many of these lessons provide an insight into how and what happens in and around professional organizations when they adopt AI. Although these lessons provide much insight, questions remain: So what? Given the likelihood of the developments described, what can we do as an organization? How do we plan to respond? Therefore in this section I provide implications for managers and decision-makers within these organizations.

Before we dive into the read thread behind the implications, several guiding principles need to be mentioned. These guiding principles are principles that decision-makers should keep in mind anyway when adopting and using AI.

- First, take into consideration that pre-existing tacit knowledge should be valued and remains important. Probably, tacit knowledge' value will only grow as a source for high quality AI models, as countervailing power, and as a valuable source of knowledge in itself.
- Second, take into consideration that realizing public values is important (Moore, 1995), and that AI adoption implies coping with multiple values. Innovating with AI must be done with awareness of values and ethical principles. There are several examples to date where incorporation of key values has fallen short. This requires a very delicate balance, where the need not to fall behind should not mean that values and the ethics of AI are overlooked.
- Third, understand that to deal with a variety of values a variety of professionals will compete for attention. Discretion will most probably be re-distributed and new dependencies will emerge.
- Fourth, understand that professionals that acknowledge other professionals are key in AI adoption. After all, AI adoption AI leads to the emergence of a new group of professionals with a different background and skillset. With growing variety and the need to interact, comes growing need for mutual understanding.

Keeping the principles in mind, the main message, further elaborated below, is that public organizations are encouraged to adopt AI and innovate with AI. Not because successful AI adoption is guaranteed, but

because there is no way back and the outside world will do so anyway. Hence, abstaining from it is too risky.

Incidentally, however, there is another message to keep in mind, that provides some limits to the positive message that AI innovation should be encouraged. In this regard, this dissertation refers to Van Der Steen (2024) who discusses *watchful waiting* in the context of AI adoption. The concept refers to actively seeking and monitoring developments and signals around AI. Without become passive, it is about being cautious, because it is hard to predict which stage technologies are working towards and which stage we are currently in. Cautiousness is necessary, because public organizations cannot risk becoming too dependent on external provides of AI technologies, and are expected to fulfil important public values which may be compromised by improper AI adoption and use. Public organizations are expected to use AI ethically. This dissertation concurs with the words by Van der Steen (2024), providing organizational suggestions on what active but cautious engagement with AI adoption might look like.

Such cautiousness is particularly important given growing public GenAI experimentation and adoption worldwide (Bick et al., 2025; European Economic and Social Committee, 2025; TNO, 2025). Many GenAI tools and applications are owned and develop by American Big Tech companies. From a European perspective, enthusiastic adoption of these solutions risks strategic dependency. While many solutions to this strategic dependence lie beyond individual organizations – because of the all-encompassing and cross-border nature of (Gen)AI - they should recognize this challenge and pursue AI sovereignty, as recommended by Adviesraad Internationale Vraagstukken (2025). This requires advocating for European or even more 'local' AI models and infrastructure while attracting and developing talent capable of building European and local models and applications.

Beyond these strategic considerations, public organizations face immediate practical challenges in AI adoption. This has been mostly the focus of this dissertation. The recommendations below follow from the thought that in the current situation organizations need intermediaries, i.e. boundary spanner, who develop a professional skillset as such. However, over time, organizations ideally may move towards an organization in which boundary spanners are no longer necessary, wherein boundary spanning is internalized. So how do organizations and decision-makers encourage AI innovation? Table 16 outlines the main recommendations that I will discuss.

Table 16: Overview of recommendations

Overarching recommendation	Practical recommendations
Recommendation #1: Facilitate novel AI professionalism	1. Give room to emerging boundary spanners 2. Develop a high tolerance for variety and overlap and embrace flexibility 3. Educate all professionals to become AI-able
Recommendation #2: Stimulate experimentation and interaction	1. Design for learning by doing, tolerate failure 2. Establish dedicated innovation time in organizational processes 3. Design for mutual engagement between incumbent and novel professionals 4. Provide guardrails and define objectives loosely 5. Design for various forms of interaction
Recommendation #3: Accept imperfection and mistakes for the sake of progress	1. Accept ambiguity regarding roles, meanings, organizational structures, and goals 2. Design for ambidexterity - failure and uncertainty should not develop into chaos and disruption

7.4.1. Facilitate novel AI professionalism

We encourage organizational decision-makers and managers to facilitate novel AI professionalism. These decision-makers should allow for job

enlargement or the emergence of new types of professionals as a result of AI adoption, even if these novel professional or enlarged roles do not fit within existing organizational structures or frameworks. Practically this means:

1. Give room to emerging boundary spanners
2. Develop a high tolerance for variety and overlap and embrace flexibility
3. Educate all professionals to become AI-able

First, this dissertation vividly shows the critical role that boundary spanners played. These actors, who speak the language of different domains, were able to transfer and translate sometimes difficult concepts across domains. In doing so, they played a critical role in the AI adoption processes. Due to the fact that AI adoption is still in its infancy, these actors are currently indispensable. In addition, given that AI adoption inherently requires interaction between many different disciplines, it is not inconceivable that these boundary spanners remain critical. Therefore, ensure that professionals growing in this boundary spanning role are not constrained by existing structures, or the lack of official designations for their role. However, do make sure that they be held accountable somewhere.

Second, develop a high tolerance for variety and overlap and embrace flexibility in tasks within and across existing roles. As noted (see section 7.2.1.) AI adoption increases the required variety in types of professional roles, necessary skills, and the variety within professional roles. Decision-makers should embrace this increased variety. As chapter 4 highlights, some end-users want to be more involved in AI development, while others are a more hesitant. Organizations should offer the opportunity to end-users who wish to extend their role to include innovation-related issues, but also offer the opportunity to others not to do so. In doing so, organizational complexity increases. The organization becomes messier

with more overlap and less distinct professional boundaries and separated professions. However, tolerating variety and embracing flexibility enables the development of more broadly accepted tools, which may actually add value within the use domain.

Third, educate all professionals to become AI-able. Start to train professionals with an affinity for AI, but also with sound knowledge of the organizations operating domain. Such professionals can become pioneers for how AI can be used within the organization. However, ultimately all professionals must engage with AI, there is no opting out. Hence, provide specialized training for those with AI affinity, but also ensure everyone develops basic AI literacy relevant to their domain. Educate incumbent professionals and show that their professional autonomy may diminish as discretionary power shifts toward systems and organizational actors who develop them. Also, incorporate values into AI-related training. Make the professionals aware that making choices in AI adoption processes is making a choice between different values. For example, some types of AI models offer more transparency than others.

7.4.2. Stimulate experimentation and interaction

Secondly, I recommend to stimulate experimentation and interaction. Decision-makers may make space for actors to interact and get to work on developing and using this new technology: AI. Learning about AI and its effects is about *learning by doing*. It is about interacting with each other, for example developers and end-users, and about interacting with the technology itself, for example an ML or GenAI tool. This perspective is much in line with the literature in organization science, for example Orlikowski (2002), who highlights how knowledge evolves in practice. According to her, knowledge is developed and maintained in practice, and can hardly be separated from it. Hence, learning about AI is about practicing with AI, instead of only learning about it statically.

Practically this means:

1. Design for learning by doing, tolerate failure
2. Establish dedicated innovation time in organizational processes
3. Design for mutual engagement between incumbent and new professionals
4. Provide guardrails and define objectives loosely
5. Design for various forms of interaction

First, design for learning by doing. In the same way that professionals develop tacit knowledge about their field through putting in the hours, it is necessary to put in the hours with AI. By 'doing' professionals themselves learn what does and does not work. By letting professionals make mistakes and interact, organizations and the decision-makers within start to understand what kind of structures and roles work, and which do not. Learning by doing also means to tolerate failure. Failures are bound to happen in AI adoption. Old and new ways of working come together, and lead to potentially conflicting situations. Misunderstandings occur, for example about the needed data, or how to segment it. Such misunderstandings are a natural consequence of a novel AI-driven way of working that enters an existing organization. Hence, accept that they occur, as they do not signal failure, but they signal that AI adoption has actually started. Accept that failures are a natural part of the process and understand that they may involve costs now, they will leads to benefits later on

Second, establish dedicated innovation time in organizational processes. Effective AI innovation requires time to research the problem, analyze possibilities, and to interact between different organizational actors. By the nature of their job, developers may dedicate much time to innovation, however for intended end-users this may be much harder. Next to having a innovation department with people naturally dedicated to innovation, organizations should give motivated end-users dedicated time to work on AI-related innovation. This upfront time investment may leads to benefits

later on. The same arguments apply to providing financial and technical resources.

Third, design for mutual engagement between incumbent and new professionals. Stimulate these professionals to mutually engage with each other and peek into each other operations. Exemplary are the dynamics between use-domain oriented end-users and more technically-oriented developers. Some end-users struggle to understand how AI might help them in the future. Additionally, they find it odd that developers or data scientists start to build AI models, while they have zero understanding of 'the domain'. By mutual engagement, developers engaging with the daily practice of the end-users and end-users with the daily practice of developers, knowledge can be shared bidirectionally, potentially creating a shared understanding of how the two worlds overlap and where one can help the other.

Fourth, provide guardrails and define objectives loosely. As shown throughout this dissertation, AI innovation does not tend to happen without direction. Without direction, it is difficult to justify why professionals need to free up time and space. Additionally, roaming without a purpose is inefficient from a managerial perspective, as it may lead to the investment of precious resources without any clear returns. On the other hand, organizations must also create space for experimentation, goal-seeking behaviour, and unexpected discoveries that emerge during the AI development process. Hence, by providing guardrails and loosely defining innovation objectives professionals get just enough direction, but also just enough autonomy to negotiate the specifics and how to get there. Examples of guardrails may be a fixed set of time to be dedicated a particular innovation project or the obligation to produce a concrete deliverable. Through this a sense of shared ownership tends to appear. That shared ownership potentially brings actors with different backgrounds closer together and makes AI innovation more successful.

Finally, design for various forms of interaction. As seen throughout this dissertation, AI adoption necessitates interaction. However, interaction can be organized in many ways. This dissertation highlighted a dedicated co-creation project in which internal developers and end-users interacted, and a more fluid and flexible interaction process facilitated by an external developer team. While outcomes varied, the point is that there is no single interaction approach for AI adoption. Some types of organized interaction work in one organization, while others work in other settings. Hence, organizations are encouraged to design for various forms of interaction, from low-cost interaction at the coffee machine, to high-cost and high-intensity co-creation programs seen throughout the dissertation.

7.4.3. Accept imperfection and ambiguity for the sake of progress

Finally, I recommend to accept imperfection and ambiguity for the sake of progress. Decision-makers are encouraged to accept that innovation comes with imperfection. This imperfection and ambiguity are part of the change process towards and organization that incorporates AI into its operations. Practically this means:

1. Accept ambiguity regarding roles, meanings, organizational structures, and goals
2. Design for ambidexterity - failure and uncertainty should not develop into chaos and disruption

First, accept ambiguity regarding roles, meanings, organizational structures, and goals. Accept that the roles of emerging professionals are not materialized yet, and it is unclear in which department up and coming boundary spanners currently 'belong'. Accept that incidentally goals from one side of the organization are not in line with the goals in other parts of the organization. Realize that this incongruity even has advantages, as it secures different values within the organization. Accept that misunderstandings between actors with different backgrounds will

occur. Over time, expect such ambiguities to decrease, as the novel terms, structures, and roles carve out a place in the organization.

Second, design for ambidexterity – failure and uncertainty should not develop into chaos and disruption. As seen, many of the recommendations are about tolerating uncertainty and failure. This creates a dilemma. On the one hand, uncertainty and failure are inherent to AI-driven innovations. On the other hand, too much uncertainty and failure can lead to chaos, disruption of the professional process and ultimately a distrust of innovation. Hence, organizational arrangements are necessary that facilitate this dilemma. Public organizations could exploit the ambidextrous organizational structure (O'Reilly & Tushman, 2004), meaning that they create organizationally distinct units, some focusing on exploitation and others on exploration, while tightly integrating these units at the senior executive level. For example, organizations could create an innovation team or department that revolves around data science or innovation, especially in the early phases of AI adoption. However, as this dissertation highlights for successful AI adoption interactions between many actors is necessary. Hence, over time drawing these organizational units closer together or findings alternative arrangements may become a necessity.

7.5. Contributions to the literature

All the empirical findings in this dissertation – next to their intrinsic value - together add to the literature. Here I highlight two main contributions to the main areas of interest: professionalism in motion, and blurring and emerging boundaries.

7.5.1. Professionalism in motion

This dissertation highlights that professionalism is in motion with public AI adoption. The classical approach to the professional, say a doctor,

with the rest of the organisation subservient this doctor, is no longer appropriate in the age of AI. Today, many professionals exist and emerge because of AI adoption. That professionalism is in motion is highlighted according to the following contributions.

1. First, this dissertation adds the notion of *re-distributed discretion.* In the professional landscape where AI has a place, a redistribution of discretion takes place, for example from end-users to novel actors such as developers or privacy specialists. Decision space moves from some actors, to others. With that, this dissertation adds to a mix of literature (Lorenz et al., 2022; Young et al., 2019), that mentions discretion in the world of AI. It highlights that discretion is not only a questions of a user and the AI model, but that discretion in the world of AI can be understood in a broader and more dynamic manner.
2. Second, this dissertation contributes to the small but growing debate on the relationship between tacit knowledge and AI. Although some highlight that tacit knowledge may be embodied in AI models (Brynjolfsson et al., 2023), thereby partly decreasing tacit knowledge importance, this dissertation argues that the relationship between AI and tacit knowledge may turn out to be quite different. It highlights the necessity of tacit knowledge as source of AI, or as a countervailing power. In doing so, it theorizes about relationships an gives an initial overview of scenarios that might occur.
3. Third, this dissertation reveals the increased variation that could be expected in the age of AI. In organizations that rely more and more on AI, it is very plausible that there is an increasing variety of different types of professionals dealing with an increasingly varied types of tasks.

7.5.2. The multiple effects of interaction

This dissertation offers a glimpse in to the various effects that interaction for the sake of AI adoption has on public organizations and the AI models itself. In doing so, it contributes to – again – a mixed set of literature spanning multiple fields (Lorenz et al., 2022; van der Voort et al., 2019; Waardenburg & Huysman, 2022). This mix of contributions shows once again that the impact of AI on public organizations is an issue that needs to be looked at from multiple angles. The contributions are:

1. First, this dissertation adds to the literature on interaction for AI adoption and demonstrates how interaction not only has the hoped-for effects of blurring boundaries, but also leads to the emergence of novel boundaries. This phenomenon – highlighted as *the waterbed effect of co-creation* in chapter 4 – adds to this area of interest as it gives novel understanding about the effects of interaction for AI adoption.
2. Second, this dissertation demonstrates the growing importance of boundary spanners throughout AI adoption processes. It demonstrates the critical role that these types of actors have by translating and transforming knowledge between different types of professionals. Although the literature also points out that actors that acts as boundary spanners may unintentionally draw power to themselves in critical processes (Waardenburg et al., 2022), it seems critical in the current landscape that boundary spanners are given space. As such, this findings adds to this area of interest as it points towards a critical factor in the success of AI adoption in public organizations.
3. Last, although less mentioned, this dissertation adds to the literature that touches upon interaction as a means to an end, in this case interaction to support the explainability of AI models (de Bruijn et al., 2021). This dissertation gives rise to a reconsideration of how explainability may be understood, and

what that means for the public organization. It highlights that interaction may be a means to achieving more explainability. Not in terms of the model, but in terms of the processes behind what AI adoption means for the organizational actors and how their professional might change. As such, this thesis gives rise to a reconsideration of what explainability means in terms of AI, and how interaction plays a role.

7.6. Future research

Chapters 4, 5, and 6 each have presented detailed directions for future research in line with contributions in the particular fields. In this section we integrate the discussed research directions, and highlight overarching future research directions that arise from this dissertation.

7.6.1. Understanding interaction-consequences relationships

Mostly in chapter 4, but also throughout chapter 5, we have explored some form of interaction within AI development processes. The dissertation shows that interaction leads to various advantages and disadvantages and can lead to surprising things, including the emergence of novel boundaries. While this dissertation sheds light on the consequences of interaction, it provides less explanations for the relationships between interaction and its consequences.

Future research may investigate specific relationships between forms of interaction and the consequences such interaction has. Does co-creating AI systems in a certain way always lead to specific consequences, and what about different contexts? As we have seen AI adoption is partly dependent on the context of organizations. Do different context leads to different outcomes in interaction AI adoption processes?

7.6.2. Investigating questions of power and transaction costs in AI adoption

Most explicitly chapter 4, but also chapter 5, has highlighted that AI adoption increasingly leads to demands on organizational stakeholders originally not involved in AI-related tasks. As a results of decisions 'higher up' in the organization, actors like inspectors or project managers are more or less compelled to contribute to AI adoption.

However, as this dissertation has highlighted, often these actors are not interested, for various reasons. On such reason may be the existence of pragmatic boundaries. In that instance, the goals and interest of involved actors are considerable different. It takes much effort to overcome such boundaries. As highlighted, the existence of transactions costs may be one reason for the occurrence of pragmatic boundaries in AI innovation processes. However, besides our notion that transaction cost play a part, we have not yet researched this further.

Hence, future research could take a more 'economical lens' and focus on the distribution of transactions casts among involved actors. Which costs do these stakeholders experience or perceive when they are becoming or not becoming involved in AI innovation processes. How is the distribution of transaction costs among these actors determined by their work processes and other legacies?

7.6.3. Exploring effects on professionals

Most explicitly in chapter 6, but also throughout chapter 4 and 5, we have explored the effect that AI adoption has on professionals and how these professionals interact in AI development processes. We have argued that the professionals landscape is changing. For example, that variety increases in terms of types of professionals, the number and types of tasks, and within professional roles. Also that boundary spanning seems to become a profession in itself in the age of AI adoption. However, chapter 4 and 5 also give clear examples of organizational actors that are

reluctant to participate in the novel development that AI adoptions asks of them. The questions is, why is that the case? And what happens within professionals when AI enters their daily lives.

Without make normative statements about the importance of contributing or not contributing to AI adoption and the value of AI adoption itself, future research could focus on the (psychological) mechanisms that are at the core of professionals in their transformation into professionals that work with AI. What are their hindrances that determine whether they do or do not want to contribute? What are the explanations for resistance in the AI adoption process and what can we learn from that resistance?

Additionally, specifically chapter 6 has highlighted the perspectives that professionals have with regard to the emergence of GenAI, and what that means for their professional work. Because our research has mostly asked about perspectives, future research could focus on experiments that uncover deeper mechanisms that explain the relationship between tacit knowledge and (generative) AI.

7.7. Personal reflections on research process: Exploratory research is difficult!

After doing qualitative exploratory research for several years, a personal reflection on the research process is in order. The reflection leads me to conclude that exploratory research is difficult and remains to be so, for many reasons. As an exploratory researcher focused on AI in public organizations – especially at the beginning of the trajectory – I felt much uncertainty.

That uncertainty relates to a lot of different aspects. Just to name a few that stood out for me: 1) To what extent do I use theory and to what extent

do I let the empirical material lead? 2) Whom or which organizations should I approach to get data? 3) What position do I take towards my data subjects? 4) Can I actually claim this based on the data I have? I have come to believe that asking such questions is innate to exploratory research and may even signal that you are doing exploratory research in 'the right way'. It seems as if exploratory research should be uncertain and uncomfortable. Throughout that process the exploratory researcher more and more learns what do to in which situation.

Over the past few years, I have learned that there are some practices or principles that might help the exploratory researcher throughout their work. In any case, they helped me. Without wanting to argue that other exploratory researches should do the same, these principles are:

- First, *work with an open mindset.* For me, this is the most important principle, but also the most difficult one to adhere to. As people, we like to make a coherent story out of signals we receive. In research, when the data seems to point one way, we – or at least me- like to draw conclusions. Within exploratory research, however, it is essential to keep an open mind. During collecting of new data, or a discussion with a peer or co-author the exploratory researcher should be open to updating the conclusion, because novel data or insights point into that direction.
- Second, *embrace the uncertainty.* Like the first principle embracing uncertainty is not easy. However, it is also critical. Especially at the beginning of the research process, or at the start of data collection for a novel research idea, there is much uncertainty. The exploratory researcher is unaware about where the research will lead. However, this uncertainty is innate to exploratory research.
- Third, *be practical.* Within the research process, you will sometimes have to make choices because they are more

practical than others. For example, initially you wanted to research a particular organization, but an opportunity presents itself at a different organization. In such a situation, it usually helps to be practical, and choose where you can actually get started. That this requires an adjustment in what you are exploring has to be taken at face value.

- Finally, *keep distance but provide value.* Do not be afraid to share research insights with the subjects you have studied. Indeed as an exploratory researcher one should try to keep some distance from the subjects. At the same time, a part of the research' value is in its practical value. Therefore, do not be afraid to share insights gained.

Sticking to these principles probably will not make exploratory research easy. However, keeping them in mind may make it less difficult.

Generative AI statement and reflection

During the preparation of this work the author used DeepL and Claude in order to: 1) perform proofreading and editing on the written manuscript, 2) brainstorm about titles and sub-titles, 3) translate pieces of text. After using this tool/service, I reviewed and edited the content as needed and take full responsibility for the content of the dissertation.

In my humble opinion, a researcher focused on AI on organizations, especially when also focusing on the effects of AI on professionals, will have to at least experiment with AI tools. Otherwise, the statements this researcher makes are only based on theory and less on practice. By only basing yourself on theory, you run the risk of misunderstanding actual effects. This researcher being myself, I indeed leveraged on the promise AI tools, in particular GenAI tools can have. While being hesitant to use them at first, I realized that I myself, as a 'professional researcher' should at least have a basic understanding of their working and experience with working with them. Given this attitude, it becomes clear that I value knowledge that is practically applicable and tested. I believe that – especially in the sometimes fuzzy world of social science - researcher should not sit in their 'ivory towers'. Instead, they should engage with practice to see whether thoughts and theories hold true in reality. Hence, this is exactly what I did. Indeed, trough personal experience I find how important it is to have a certain level of experience and knowledge, to be able to assess the output of AI tools. Without it, I am indeed unable to say or assess whether the output is of any quality. As such, findings mentioned by professionals in the field also seem to be more or less true for me as a researcher.

List of publications

van Krimpen, F. van der Voort, H. de Bruijn, J.A. Generative AI and Professionals' Tacit Knowledge: Exploring Possible Relationships. (manuscript submitted to Public Administration Quarterly)

van Krimpen, F.J., Arnaboldi, M. Querini, L. Explainable AI as a Learning Process: The Impact of an Interactive and Iterative Development Process on XAI for Organizational Stakeholders. (revised manuscript submitted to Government Information Quarterly Special Issue: AI in Government: Design, Implementation and Oversight)

Krimpen, F. van, Voort, H. van der, & Bruijn, H. de (2025). De impact van generatieve AI op toezichthouders: Een empirische verkenning onder professionals. Tijdschrift voor Toezicht, 16(2-3), 44-55. https://doi.org/10.5553/TvT/187987052025016002002

van Krimpen, F., & van der Voort, H. (2025). Co-creation with machine learning: Towards a dynamic understanding of knowledge boundaries between developers and end-users. Information and Organization, 35(2). https://doi.org/10.1016/j.infoandorg.2025.100574

van Krimpen, F., de Bruijn, H., & Arnaboldi, M. (2023). Machine learning algorithms and public decision-making: A conceptual overview. The Routledge Handbook of Public Sector Accounting, 124–138. https://doi.org/10.4324/9781003295945-12

(adapted for chapter 2 of this dissertation)

List of presentations

Conference presentations

Reference	Date of presentation
Van Krimpen, F.J., van der Voort, H. (2022). An exploration of factors that explain cognitive distance between developers and end-users of Machine Learning tools in public organizations, NIG Annual Work Conference 2022, Tilburg	13/10/2022
Van Krimpen, F.J., Arnaboldi, M., Querini, L. (2023) Navigating the Black Box: Fostering Explainable AI through Multi-Stakeholder Interactions, NIG Annual Work Conference 2023, Delft, 2/11/2023	2/11/2023
Van Krimpen, F.J., Van der Voort, H.G., De Bruijn, J.A. (2024) Generative AI and Public Professionals in Inspectorates, EGOV Conference, Leuven, 3-5 September, 3/9/2024	3/9/2024

Practitioner presentations

Reference	Date of presentation
Presentation at Human Environment and Transport Inspectorate (ILT).	*21/3/22*
Presentation at Immigration and Naturalization Service (IND)	*2/2/22*
Presentation at Human Environment and Transport Inspectorate (ILT).	*10/8/22*
Presentation at Human Environment and Transport Inspectorate (ILT).	*6/10/22*
Report for the Human Environment and Transport Inspectorate	*30/11/22*
Presentation at Human Environment and Transport Inspectorate (ILT)	*25/1/23*
Presentation at ILT Innovators Network	*21/2/23*
Participate at the yearly Oversight Festival	*18/06/24*

About the author

Floris Justus van Krimpen was born in Capelle aan den Ijssel, The Netherlands, on February 23, 1994. He holds a BSc in Systems Engineering, Policy Analysis and Management (TU Delft) obtained in 2017, and a MSc in Engineering and Policy Analysis (TU Delft) obtained in 2019. During his MSc programme, he spent an exchange semester at Politecnico di Milano. His exchange semester in Milan contributed to him starting a Dual PhD at both TU Delft and the Politecnico di Milano. Before starting a PhD he followed the L.E.A.D. Programme, a one year traineeship from X!Delft, a part of the TU Delft valorisation centre. From 2020-2025 he was a PhD Researcher at TU Delft and Politecnico di Milano. In his research he focused on the implications of the adoption of AI for public organizations, with a focus on regulatory agencies. In 2025, he started working as a management consultant and researcher at Bisnez, mostly focused on organizational change processes and digital transformation in the public sector.

Bibliography

Abrahamsson, P., Salo, O., Ronkainen, J., & Warsta, J. (2017). Agile software development methods: Review and analysis. *ArXiv Preprint ArXiv:1709.08439*.

Adviesraad Internationale Vraagstukken. (2025). *Democratische waarden in het Nederlandse buitenlandbeleid*.

Alexopoulos, C., Lachana, Z., Androutsopoulou, A., Diamantopoulou, V., Charalabidis, Y., & Loutsaris, M. A. (2019). How Machine Learning is Changing e-Government. *Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance*, *Part F1481*, 354–363. https://doi.org/10.1145/3326365.3326412

Alford, J., & Head, B. W. (2017). Wicked and less wicked problems: A typology and a contingency framework. *Policy and Society*, *36*(3), 397–413. https://doi.org/10.1080/14494035.2017.1361634

Algemene Rekenkamer. (2024). *Helft van de ouders wacht nog, maar hersteloperatie Toeslagen wel op gang*. https://www.rekenkamer.nl/actueel/nieuws/2024/05/15/helft-van-de-ouders-wacht-nog-maar-hersteloperatie-toeslagen-wel-op-gang

Alpaydin, E. (2020). *Introduction to Machine Learning* (Fourth Edi). MIT Press Ltd.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Avison, D. E., Lau, F., Myers, M. D., & Nielsen, P. A. (1999). Action research. *Communications of the ACM*, *42*(1), 94–97. https://doi.org/10.1145/291469.291479

Bailey, D. E., Faraj, S., Hinds, P. J., Leonardi, P. M., & von Krogh, G. (2022). We Are All Theorists of Technology Now: A Relational Perspective on Emerging Technology and Organizing. *Organization Science*, *33*(1), 1–18. https://doi.org/10.1287/ORSC.2021.1562

Bailey, D., Faraj, S., Hinds, P., von Krogh, G., & Leonardi, P. (2019). Special issue of organization science: Emerging technologies and organizing. *Organization Science*, *30*(3), 642–646. https://doi.org/10.1287/orsc.2019.1299

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, *104*(671), 671–732. https://doi.org/10.2139/ssrn.2477899

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*(December 2019), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Barrett, M., Oborn, E., Orlikowski, W. J., & Yates, J. A. (2012). Reconfiguring boundary relations: Robotic innovations in pharmacy work. *Organization Science*, *23*(5), 1448–1466. https://doi.org/10.1287/orsc.1100.0639

Barriball, K. L., & While, A. (1994). Collecting data using a semi-structured interview: a discussion paper. *Journal of Advanced Nursing*, *19*(2), 328–335. http://www.ncbi.nlm.nih.gov/pubmed/8188965

BBC. (2020). *A-levels: Algorithm at centre of grading crisis "unlawful" says Labour*. https://www.bbc.com/news/uk-politics-53837722

Bechky, B. A. (2003). Sharing meaning across occupational communities: The transformation of understanding on a production floor. *Organization Science*, *14*(3). https://doi.org/10.1287/orsc.14.3.312.15162

Belle, V., & Papantonis, I. (2021). Principles and Practice of Explainable Machine Learning. *Frontiers in Big Data*, *4*(July), 1–25. https://doi.org/10.3389/fdata.2021.688969

Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology*, *13*(2), 193–216. https://doi.org/10.1007/s11292-017-9286-2

Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., Ghosh, J., Puri, R., Moura, J. M. F., & Eckersley, P. (2020). Explainable machine learning in deployment. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 648–657. https://doi.org/10.1145/3351095.3375624

Bick, A., Blandin, A., & Deming, D. J. (2025). *The Rapid Adoption of Generative AI, Federal Reserve Bank of St. Louis Working Paper 2024-027*. https://doi.org/10.20955/wp.2024.027.

Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, *62*(2), 174–184. https://doi.org/10.1111/0033-3352.00168

Brasse, J., Rebecca, H., Maximilian, B., Mathias, F., & Irina, K. (2023). Explainable artificial intelligence in information systems : A review of the status quo and future research directions. *Electronic Markets*, *33*(1), 1–30. https://doi.org/10.1007/s12525-023-00644-5

Bright, J., Enock, F. E., Esnaashari, S., Francis, J., Hashem, Y., & Morgan, D. (2024). *Generative AI is already widespread in the public sector*. 1–10. http://arxiv.org/abs/2401.01291

Brown, J. S., & Duguid, P. (1991). Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. *Organization Science*, *2*(1), 40–57.

Brynjolfsson, E., Li, D., & Raymond, L. (2023). Generative Ai at Work. In *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4426942

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530–1534. https://doi.org/10.1126/science.aap8062

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data and Society*, *3*(1), 1–12. https://doi.org/10.1177/2053951715622512

Busuioc, M. (2022). AI algorithmic oversight: New frontiers in regulation. *Handbook of Regulatory Authorities*, *Geiger*, 470–486. https://doi.org/10.4337/9781839108990.00043

Campion, A., Gasco-Hernandez, M., Jankin Mikhaylov, S., & Esteve, M. (2022). Overcoming the Challenges of Collaboratively Adopting Artificial Intelligence in the Public Sector. *Social Science Computer Review*, *40*(2), 462–477. https://doi.org/10.1177/0894439320979953

Canterino, F., Shani, A. B. (Rami), Coghlan, D., & Brunelli, M. S. (2016). Collaborative Management Research as a Modality of Action Research. *The Journal of Applied Behavioral Science*, *52*(2), 157–186. https://doi.org/10.1177/0021886316641509

Carlile, P. R. (2002). A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science*, *13*(4). https://doi.org/10.1287/orsc.13.4.442.2953

Carlile, P. R. (2004). Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organization Science*, *15*(5), 555–568. https://doi.org/10.1287/orsc.1040.0094

Darsø, L. (2001). *Innovation in the Making*. Samfundslitteratur. https://books.google.nl/books?id=bBVPPgAACAAJ

de Bruijn, H. (2012). *Managing Professionals*. Routledge. https://doi.org/10.4324/9780203848272

De Bruijn, H., & Ten Heuvelhof, E. (2018). *Management in Networks* (Second Edi). Routledge - Taylor & Francis Group. https://doi.org/10.4324/9781315453019

De Bruijn, H., & Ten Heuvelhof, E. (2019). *Handhaving: Het spel tusen inspecteur en inspectee*. Boom bestuurskunde.

de Bruijn, H., Warnier, M., & Janssen, M. (2021). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*, *December*, 101666. https://doi.org/10.1016/j.giq.2021.101666

Delfos, J., Zuiderwijk, A., Van Cranenburgh, S., & Chorus, C. (2022). Perceived challenges and opportunities of machine learning applications in governmental organisations: an interview-based exploration in the Netherlands. *ACM International Conference Proceeding Series*, 82–89. https://doi.org/10.1145/3560107.3560122

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4573321

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87. https://doi.org/10.1145/2347736.2347755

Donaldson, L. (2001). The Contigency Theory of Oganizational Design: Challenges. *Organization Design*, 284.

Du, M., Liu, N., & Hu, X. (2020). Techniques for interpretable machine learning. *Communications of the ACM*, *63*(1), 68–77. https://doi.org/10.1145/3359786

Eggers, W. D., Schatsky, D., & Viechnicki, P. (2017). AI-augmented government. Using cognitive technologies to redesign public sector work. *Deloitte Center for Government Insights*, 1–24. https://www2.deloitte.com/content/dam/insights/us/articles/3832_AI-augmented-government/DUP_AI-augmented-government.pdf%0Ahttps://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/artificial-intelligence-government.html

European Economic and Social Committee. (2025). *Generative AI and foundation models in the EU : Uptake , opportunities , challenges , and a way forward*.

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, *28*(1), 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Ferns, D. C. (1991). Developments in programme management. *International Journal of Project Management*, *9*(3), 148–156. https://doi.org/10.1016/0263-7863(91)90039-X

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, *25*(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Galbraith, J. R. (1973). *Designing Complex Organizations*. Addison-Wesley Publishing Company. https://books.google.nl/books?id=FUe3AAAAIAAJ

Gerlich, M. (2025). *AI Tools in Society : Impacts on Cognitive Offloading and the Future of Critical Thinking*. 1–28.

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(30). https://doi.org/10.1073/pnas.2305016120

Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). *ChatGPT is not all you need. A State of the Art Review of large Generative AI models*. http://arxiv.org/abs/2301.04655

Greenwood, D., & Levin, M. (2007). *Introduction to Action Research*. SAGE Publications, Inc. https://doi.org/10.4135/9781412984614

Hammarberg, K., Kirkman, M., & De Lacey, S. (2016). Qualitative research methods: When to use them and how to judge them. *Human Reproduction*, *31*(3), 498–501. https://doi.org/10.1093/humrep/dev334

Head, B. W., & Alford, J. (2015). Wicked Problems: Implications for Public Policy and Management. *Administration and Society*, *47*(6), 711–739. https://doi.org/10.1177/0095399713481601

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science and Medicine*, *292*, 114523. https://doi.org/10.1016/j.socscimed.2021.114523

Holmström, J., & Hällgren, M. (2022). AI management beyond the hype: exploring the co-constitution of AI and organizational context. *AI and Society*, *37*(4), 1575–1585. https://doi.org/10.1007/s00146-021-01249-2

Hong, S. R., Hullman, J., & Bertini, E. (2020). Human Factors in Model Interpretability: Industry Practices, Challenges, and Needs. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW1), 1–27. https://doi.org/10.1145/3392878

Howells, J. (1996). Tacit knowledge. *Technology Analysis & Strategic Management*, *8*(2), 91–106. https://doi.org/10.1080/09537329608524237

Hüllmann, J. A., Kimathi, K., & Weritz, P. (2024). Large-Scale Agile Project Management in Safety-Critical Industries: A Case Study on Challenges and Solutions. *Information Systems Management*, *00*(00), 1–23. https://doi.org/10.1080/10580530.2024.2349886

Hutter, B. M. (2005). The Attractions of Risk-based Regulation: accounting for the emergence of risk ideas in regulation. *CARR Discussion Paper*, *March*, 17.

Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, *33*(3), 371–377. https://doi.org/10.1016/j.giq.2016.08.011

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kellogg, K. C., Orlikowski, W. J., & Yates, J. A. (2006). Life in the trading zone: Structuring coordination across boundaries in postbureaucratic organizations. *Organization Science*, *17*(1), 22–44. https://doi.org/10.1287/orsc.1050.0157

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, *6*(February), 100225. https://doi.org/10.1016/j.caeai.2024.100225

Koop, C. (2015). Assessing the Mandatory Accountability of Regulatory Agencies. In A. C. Bianculli, X. Fernández-i-Marín, & J. Jordana (Eds.), *Accountability and Regulatory Governance* (pp. 78–104). Palgrave Macmillan UK. https://doi.org/10.1057/9781137349583_4

Laganga, L. R. (2011). Lean service operations: Reflections and new directions for capacity expansion in outpatient clinics. *Journal of Operations Management*, *29*(5), 422–433. https://doi.org/10.1016/j.jom.2010.12.005

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence*, *296*, 103473. https://doi.org/10.1016/j.artint.2021.103473

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Lawrence, P. R., & Lorsch, J. W. (1967). *Organization and Environment: Managing Differentiation and Integration*. Division of Research, Graduate School of Business Administration, Harvard University. https://books.google.nl/books?id=zy1HAAAAMAAJ

Lebovitz, S., Levina, N., & Lifshitz-Assa, H. (2021). Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

Levina, N., & Vaast, E. (2005). The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems. *MIS Quarterly*, *29*(2), 335. https://doi.org/10.2307/25148682

Lipsky, M. (1980). *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. Russel Sage Foundation. https://doi.org/10.1177/003232928001000113

Longhurst, R. (2010). Semi-structured interviews and focus groups. In *Key Methods in Geography*. SAGE Publications Ltd. https://is.muni.cz/el/1431/jaro2015/Z0132/um/54979481/_Nicholas_Clifford__Gill_Valentine__Key_Methods_in_BookFi.org_.pdf#page=126

Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Ser, J. Del, Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., Jiang, R., Khosravi, H., Lecue, F., Malgieri, G., Páez, A., Samek, W., Schneider, J., Speith, T., & Stumpf, S. (2024). Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*, *106*(February), 102301. https://doi.org/10.1016/j.inffus.2024.102301

Lorenz, L. (2024). *The Myth of Algorithmic Regulation*.

Lorenz, L., Erp, J. Van, & Meijer, A. (2022). Machine-learning algorithms in regulatory practice agencies. *Technology & Regulation*, 1–11. https://doi.org/10.26116/techreg.2022.001

Mayer, A. S., van den Broek, E., Karacic, T., Hidalgo, M., & Huysman, M. (2023). Managing Collaborative Development of Artificial Intelligence: Lessons from the Field. *Proceedings of the 56th Hawaii International Conference on System Sciences*, *4*, 6139–6148.

Mehr, H. (2017). Artificial Intelligence for Citizen Services and Government. *Harvard Ash Center Technology & Democracy*, *August*, 1–16. https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

Meijer, A., Lorenz, L., & Wessels, M. (2021). Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems. *Public Administration Review*, *81*(5), 837–846. https://doi.org/10.1111/puar.13391

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management*, *39*(1), 53–63. https://doi.org/10.1080/10580530.2020.1849465

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Ministerie van Binnenlandse Zaken en Overheidsrelaties. (2024). *Overheidsbrede Visie Generatieve AI*.

Mintzberg, H. (1979). *The Structuring of Organizations: A Synthesis of the Research*. Prentice-Hall. https://books.google.nl/books?id=NQ1HAAAAMAAJ

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill.

Moore, M. (1995). *Creating Public Value - Strategic Management in Government*. Harvard University Press.

Nonaka, I., & von Krogh, G. (2009). Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory. *Organization Science*, *20*(3), 635–652. http://www.jstor.org/stable/25614679

Noordegraaf, M. (2007). From "Pure" to "Hybrid" Professionalism. *Administration & Society*, *39*(6), 761–785. https://doi.org/10.1177/0095399707304434

O'Reilly, C. A., & Tushman, M. L. (2004). The ambidextrous organization. *Harvard Business Review*, *82*(4), 74–83.

Orlikowski, W. J. (2002). Knowing in Practice: Enacting a Collective Capability in Distributed Organizing. *Organization Science*, *13*(3), 249–273. https://doi.org/10.1287/orsc.13.3.249.2776

Ørngreen, R., & Levinsen, K. (2017). Workshops as a research methodology. *Electronic Journal of E-Learning*, *15*, 70–81.

Orr, J. E. (1996). *Talking about Machines: An Ethnography of a Modern Job*. Cornell University Press. http://www.jstor.org/stable/10.7591/j.ctt1hhfnkz

Pellegrinelli, S. (1997). Programme management: Organising project-based change. *International Journal of Project Management*, *15*(3), 141–149. https://doi.org/10.1016/S0263-7863(96)00063-4

Polanyi, M. (1966). *The Tacit Dimension*. University of Chicago Press.

Preece, A., Harborne, D., Braines, D., Tomsett, R., & Chakraborty, S. (2018). *Stakeholders in Explainable AI*. http://arxiv.org/abs/1810.00184

Prevaas, B. (2018). *Werken aan Programma's*. https://www.werkenaanprogrammas.nl/

Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. *Qualitative Research in Accounting and Management*, *8*(3), 238–264. https://doi.org/10.1108/11766091111162070

Rapoport, R. N. (1970). Three Dilemmas in Action Research. *Human Relations*, *23*(6), 499–513. https://doi.org/10.1177/001872677002300601

Richthofen, G. von, Ogolla, S., & Send, H. (2022). Adopting AI in the Context of Knowledge Work: Empirical Insights from German Organizations. *Information (Switzerland)*, *13*(4), 1–16. https://doi.org/10.3390/info13040199

Ritala, P., Ruokonen, M., & Ramaul, L. (2023). Transforming boundaries: how does ChatGPT change knowledge work? *Journal of Business Strategy*, *210166*. https://doi.org/10.1108/JBS-05-2023-0094

Russel, S., & Norvig, P. (2016). *Artificial Intelligence: A modern Approach* (3rd editio). Pearson Education Limited.

Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *44*(1.2), 206–226. https://doi.org/10.1147/rd.441.0206

Schneider, J., Abraham, R., Meske, C., & Vom Brocke, J. (2023). Artificial Intelligence Governance For Businesses. *Information Systems Management*, *40*(3), 229–249. https://doi.org/10.1080/10580530.2022.2085825

Selten, F., Robeer, M., & Grimmelikhuijsen, S. (2023). 'Just like I thought': Street-level bureaucrats trust AI recommendations if they confirm their professional judgment. *Public Administration Review*, *83*(2), 263–278. https://doi.org/10.1111/puar.13602

Siggelkow, N. (2007). Persuasion with case studies. *Academy of Management Journal*, *50*(1), 20–24. https://doi.org/10.5465/AMJ.2007.24160882

Sousa, W. G. de, Melo, E. R. P. de, Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly*, *36*(4), 101392. https://doi.org/10.1016/j.giq.2019.07.004

Strauss, S. (2023). Some Emerging Hypotheses about Using Generative AI in Public Sector Operations. *SSRN Electronic Journal*, 1–17. https://doi.org/10.2139/ssrn.4544943

Tecuci, G. (2012). Artificial intelligence. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(2), 168–180. https://doi.org/10.1002/wics.200

TNO. (2025). *Overheidsbrede Monitor Generatieve AI*. *December*.

Toezine. (2021). *Zo kan AI ook toezichthouders zélf dienen*. https://www.toezine.nl/zo-kan-ai-ook-toezichthouders-zelf-dienen/

Tomsett, R., Braines, D., Harborne, D., Preece, A., & Chakraborty, S. (2018). *Interpretable to Whom? A Role-based Model for Analyzing Interpretable Machine Learning Systems*. https://doi.org/10.48550/arXiv.1806.07552

Toosi, A., Bottino, A. G., Saboury, B., Siegel, E., & Rahmim, A. (2021). A Brief History of AI: How to Prevent Another Winter (A Critical Review). *PET Clinics*, *16*(4), 449–469. https://doi.org/10.1093/mind/LIX.236.433

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460. https://doi.org/10.1093/mind/LIX.236.433

Ulbricht, L., & Yeung, K. (2022). Algorithmic regulation: A maturing concept for investigating regulation of and through algorithms. *Regulation and Governance*, *16*(1), 3–22. https://doi.org/10.1111/rego.12437

van den Broek, E., Sergeeva, A., & Huysman Vrije, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly*, *45*(3), 1557–1580. https://doi.org/10.25300/MISQ/2021/16559

Van Der Steen, M. (2024). *Technologie in de tussentijd: AI, publieke waarde en democratie*.

van der Voort, H. G., Klievink, A. J., Arnaboldi, M., & Meijer, A. J. (2019). Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making? *Government Information Quarterly*, *36*(1), 27–38. https://doi.org/10.1016/j.giq.2018.10.011

van der Voort, H., van Bulderen, S., Cunningham, S., & Janssen, M. (2021). Data science as knowledge creation a framework for synergies between data analysts and domain professionals. *Technological Forecasting and Social Change*, *173*(August), 121160. https://doi.org/10.1016/j.techfore.2021.121160

van Erp, J., & van Wingerde, K. (2013). De responsieve toezichthouder. *Tijdschrift Voor Toezicht*, *4*(4), 26–32. https://doi.org/10.5553/tvt/187987052013004004005

Van Huffelen, A. (2023). *Kamerbrief over voorlopig standpunt voor Rijksorganisaties bij het gebruik van generatieve AI*.

van Krimpen, F., de Bruijn, H., & Arnaboldi, M. (2023). Machine learning algorithms and public decision-making: A conceptual overview. *The Routledge Handbook of Public Sector Accounting*, 124–138. https://doi.org/10.4324/9781003295945-12

Van Maanen, J., & Barley, S. R. (1984). Occupational communities: Culture and control in organizations. In *Research in Organizational Behavior* (Vol. 6, pp. 287–365). JAI Press, Inc.

Veale, M., & Brass, I. (2019). Administration by Algorithm? Public Management meets Public Sector Machine Learning. In K. Yeung & M. Lodge (Eds.), *Algorithmic Regulation* (pp. 1–30). Oxford University Press. https://doi.org/10.31235/osf.io/mwhnb

Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3174014

Viechnicki, P., & Eggers, W. D. (2017). *How much time and money can AI save government ? Cognitive technologies could free up hundreds of millions of*. 23.

Waardenburg, L., & Huysman, M. (2022). From coexistence to co-creation: Blurring boundaries in the age of AI. *Information and Organization*, *32*(4), 100432. https://doi.org/10.1016/j.infoandorg.2022.100432

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the Land of the Blind, the One-Eyed Man Is King: Knowledge Brokerage in the Age of Learning Algorithms. *Organization Science*, *33*(1), 59–82. https://doi.org/10.1287/orsc.2021.1544

Waardenburg, L., Sergeeva, A., & Huysman, M. (2018). Hotspots and blind spots: A case of predictive policing in practice. In *IFIP Advances in Information and Communication Technology* (Vol. 543). Springer International Publishing. https://doi.org/10.1007/978-3-030-04091-8_8

Weick, K. E. (1979). *The Social Psychology of Organizing* (Issue Second Edition). Random House. https://doi.org/10.3917/mana.182.0189

Williams, M., & Moser, T. (2019). The art of coding and thematic exploration in qualitative research. *International Management Review*, *15*(1), 45–55.

Williams, P. (2013). We are all boundary spanners now? *International Journal of Public Sector Management*, *26*(1), 17–32. https://doi.org/10.1108/09513551311293417

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector—Applications and Challenges. *International Journal of Public Administration*, *42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation and Governance*, *12*(4), 505–523. https://doi.org/10.1111/rego.12158

Yin, R. K. (2013). *Case Study Research: Design and Methods* (5th ed.). SAGE Publications Inc.

Young, M. M., Bullock, J. B., & Lecy, J. D. (2019). Artificial Discretion as a Tool of Governance: A Framework for Understanding the Impact of Artificial Intelligence on Public Administration. *Perspectives on Public Management and Governance*, 301–313. https://doi.org/10.1093/ppmgov/gvz014

Zuiderwijk, A., Chen, Y., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance : A systematic literature review and a research agenda. *Government Information Quarterly*, *May 2020*, 101577. https://doi.org/10.1016/j.giq.2021.101577

Zweig, K. A., Wenzelburger, G., & Krafft, T. D. (2018). On Chances and Risks of Security Related Algorithmic Decision Making Systems. *European Journal for Security Research*, *3*(2), 181–203. https://doi.org/10.1007/s41125-018-0031-2

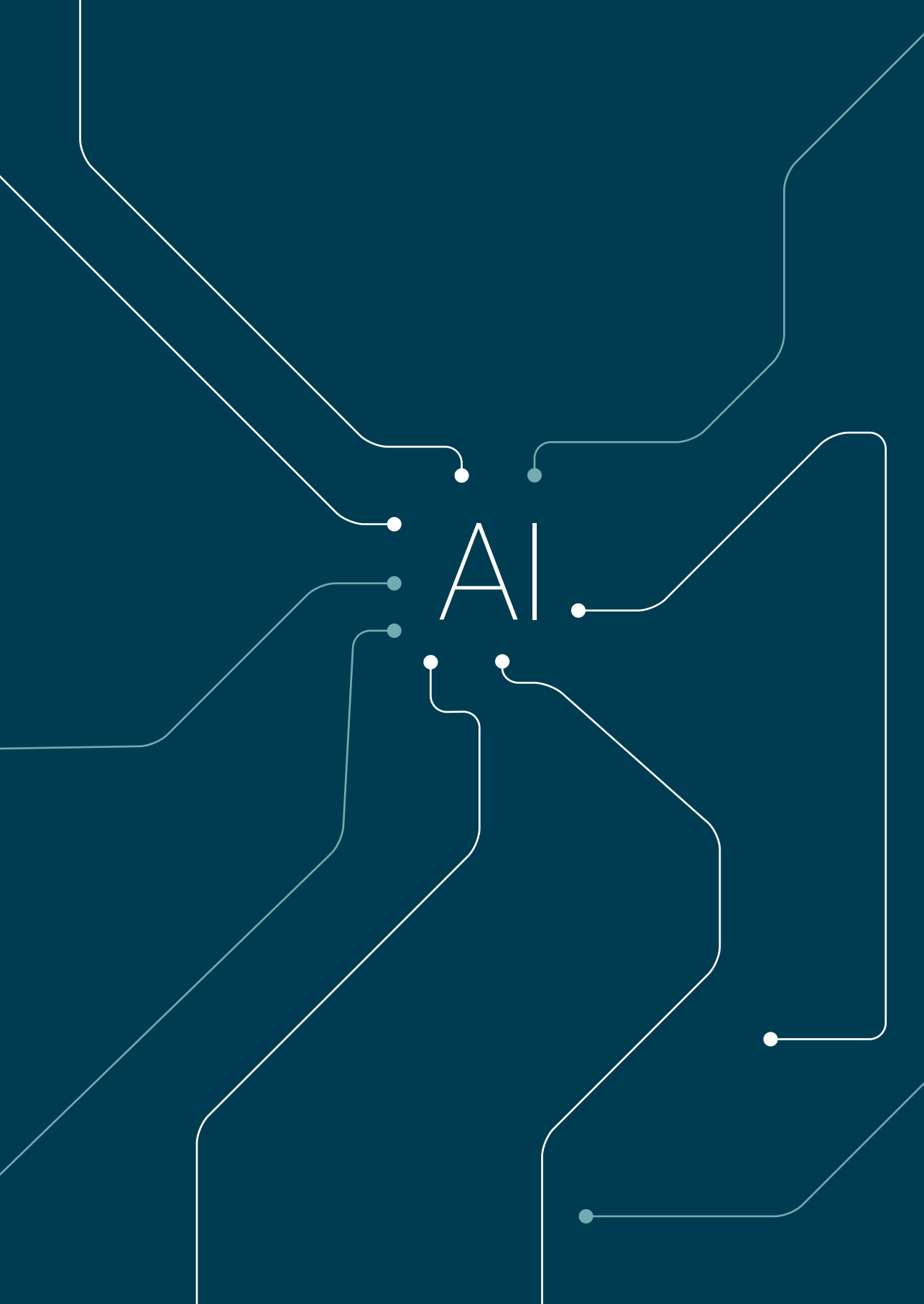
Appendix

Appendix A

Participants at meetings and interviews

Table 17: Overview of meetings/interviews and participants

Data type	***Participants***
Kick-off meeting	two middle managers from Check (heads of sub-department), seven Check managers with various duties, manager, external consultant, two process integrators, Innovators project manager
Seven semi-structured interviews	Sub-department head, deputy head, regional inspection head, business intelligence manager, innovation manager, application coordinator, external consultant, process integrator, relationship manager, project integrator, architecture representative, privacy officer
Two meetings where the results of the initial analysis were presented	Project manager at Check, department deputy head
Two internal project team meetings	Project manager at Check, data scientist at Check
Five meetings with Check	All stakeholders mentioned so far (across all meetings) plus three external consultants, one additional data scientist
One semi-structured interview	Project manager at Innovators, data scientist at Innovators


AI