Trust in AI

Designing a Practical Toolkit for Public Sector Service Development

Thesis Information

University	Aalborg University
Master Programme	Service Systems Design
Title	Trust in AI: Designing a Practical Toolkit for Public Sector Service Development
Theme	Master Thesis (30 ECTS)
Project Period	Spring 2025
Company Collaborator	Citizen Service Development (Borgerservice Udvikling)
Academic Supervisor	Nicola Morelli
Company Supervisor	Sebastian Campion
Page Numbers	112
Date of Completion	26th May, 2025
Author	Janhvi Jayesh Sampat Study Number: 20231026

AALBORG
UNIVERSITY

Abstract

We are witnessing a steady rise in the adoption of *artificial intelligence (AI)* within public services. While AI promises greater efficiency and enhanced decision-making, its opaque logic and potential for bias raise concerns about transparency, accountability, and public acceptance. This thesis investigates how public sector service creators in Copenhagen Municipality's Department for *Citizen Service Development (CSD)* might design AI-enabled services that foster trust and align with the needs of end users. Although existing research discusses the importance of trust in socio-technical systems, a gap in practical, design-oriented tools that address trust concerns in real-world service development is observed. To respond to this gap, the research explored trust as a dynamic concept shaped by cognitive, social, and institutional factors. A systemic design approach was applied to develop the Trust Toolkit intended to help public sector teams reflect on, evaluate, and integrate trust considerations throughout the development process of AI-enabled services. The toolkit was developed in collaboration with Copenhagen Municipality's Department for Citizen Service Development and shaped through co-creation and testing. While the findings suggest that practical design support for trust is both necessary and feasible, it also highlights that a toolkit is not a conclusive solution and should be critically evaluated in future work. Overall, this project contributes a practical artefact and design perspective to ongoing discussions about trust, AI, and public service innovation, while also calling for continued interrogation of how trust is conceptualised, operationalised, and maintained in public sector innovation.

Keywords: AI, trust, public sector, systemic design, design toolkit, service design

Preface

This thesis was written in Spring 2025 as part of the Master's programme in Service Systems Design at Aalborg University, Copenhagen. It marks the culmination of work undertaken during the semesters and builds on earlier academic and practical engagements in the field. The project was supervised by Professor Nicola Morelli (Department of Architecture, Design and Media Technology, AAU), and carried out in collaboration with the Department of Citizen Service Development in Copenhagen Municipality.

I would like to extend my sincere thanks to my supervisor, Professor Nicola Morelli, for his steady guidance and encouragement throughout this process. His belief in the value of my work has been both validating and motivating. I am grateful for his considered feedback and trust in my abilities. I am also deeply grateful to my mentor at CSD, Sebastian Campion. Working under his tutelage has been one of the most formative aspects of this journey. His thoughtful critique and deep knowledge about service design have shaped not just this project, but also how I see the role of design. I have learnt a great deal from him - lessons that will stay with me well beyond this thesis. I also extend my heartfelt gratitude to my manager at CSD, Karen Westman Hertz, whose trust and backing were essential to realising this project.

I would like to thank all participants involved in the observations, interviews, and workshops for generously sharing their time and insights. This project could not have taken shape without their contributions.

To my dear friends Tone Eide Hilmen, Sara Werge, and Priyal Gala - thank you for your thoughtful input, steady support, and for making the process lighter with your presence. Special thanks to Marie Jakobsen for her thoughtful guidance at key moments. I also extend my thanks to the SSD faculty for their ongoing encouragement and academic guidance throughout the course of my studies.

Finally, I am deeply thankful to my family and friends, both in Copenhagen and back home. To my parents Sonal and Jayesh Sampat, and my in-laws, Priti Ravi and Ravi Ramakantan - thank you for your love, encouragement, and belief in me across continents. To my brothers Harsh and Mayank, for cheering me on throughout. And finally, to my partner through it all, Prithvi Ravi Kantan, thank you for walking every step of this journey with me. You are my rock.

Contents

Reading Guide

This reading guide outlines the structure of this thesis, offering a brief overview of each section to help the reader navigate the content and understand how the research unfolds.

1

INTRODUCTION

The introductory chapter establishes the problem area, presents the research question, which defines the principal focus of the study, and articulates both personal and academic learning objectives.

RELATED RESEARCH

This chapter provides a review of relevant literature, including the foundational concepts of Artificial Intelligence, its application within public sector services, associated challenges, and the critical role of trust in AI. It further situates the study within the Danish socio-cultural context regarding trust and technology.

2

3

METHODOLOGY

This chapter explains the methodological approach, detailing the application of systemic design principles within the framework of the Double Diamond model, which guides the design process.

DESIGN CASE

The design case is introduced herein, offering contextual information on Copenhagen Municipality, the Department of Citizen Service Development (CSD), and the collaboratively developed project brief.

4

DESIGN PROCESS

5

This chapter documents the design process, structured according to the four phases of the double diamond: Discover, Define, Develop, and Deliver, with each phase describing specific methods, activities, and outcomes.

DISCUSSION

6

The discussion critically analyses the findings in relation to the research question, addresses encountered challenges, and reflects on the broader implications of the study.

CONCLUSION

7

The concluding chapter summarises the principal insights, contributions, and offers recommendations for future research.

1

INTRODUCTION

1.1 Introduction

Artificial Intelligence (AI) is no longer an emerging technology; it is a fundamental part of everyday life. From personalised recommendations to automated decision-making systems, AI technologies are shaping how we access information, receive services, and interact with the world (Duan et al., 2019; Huang & Rust, 2018; Kim et al., 2021; Li et al., 2021). The public sector has also begun adopting AI tools to improve operational efficiency, enhance citizen services, and manage increasing complexity. Tasks such as benefit allocation, case handling, and fraud detection are increasingly being delegated to algorithmic systems (Berryhill et al., 2019; De Sousa et al., 2019; Kuziemski & Misuraca, 2020; Mikhaylov et al., 2018; Wirtz et al., 2019). These applications promise to streamline bureaucratic processes, reduce human error, and make service delivery more consistent and efficient (Wirtz et al., 2019).

However, the integration of AI into public services introduces new ethical, technical, and societal challenges. Public institutions are held to high standards of accountability, transparency, and fairness, and any technology incorporated within these institutions must uphold these principles (Lyrio et al., 2018; Romzek, 2000; Veale et al., 2018). AI systems, by contrast, often operate opaquely, raise concerns about bias and surveillance, and can displace the human judgement that citizens may expect from public servants (Ferrara, 2023; Mensah, 2023). As a result, trust in AI has emerged as a central issue (Choung et al., 2023; Glikson & Woolley, 2020; Habbal et al., 2024; Jacovi et al., 2021; Lockey et al., 2021). Trust determines whether citizens will accept and adopt AI-enabled services, and whether institutions can maintain legitimacy in the face of technological change.

Unlike traditional technologies, AI operates with a degree of autonomy and adaptability that challenges existing social bonds. As such, the adoption of AI in sensitive public contexts must be approached carefully. Without trust, even well-designed AI systems may be met with scepticism or rejection, resulting in wasted resources and reputational damage for public bodies (Hang & Chen, 2022; Park et al., 2022). Beyond adoption, trust also influences how people interact with AI systems, how errors are perceived, and how responsibility is assigned. In this sense, trust is not merely a desirable quality but a precondition for the ethical and effective use of AI in the public sector.

The literature on trust in AI spans a range of disciplines, including computer science,

psychology, ethics, and public administration (Glikson & Woolley, 2020; M. Ryan, 2020). One strand focuses on the types of AI used in public services and the specific risks they present, such as lack of transparency, data privacy concerns, and algorithmic bias (De Fine Licht & De Fine Licht, 2020; Kuziemski & Misuraca, 2020; Mazurek & Małagocka, 2019). Another strand examines AI-enabled services in the public sector, noting the unique accountability challenges that arise when compared to private sector applications (Medeiros, 2020). A significant body of research has explored how trust is conceptualised in relation to technology. Commonly cited dimensions include interpersonal trust (between people), institutional trust (in the organisations deploying AI), and technological trust (in the performance and reliability of AI itself) (Glikson & Woolley, 2020; Hasija & Esper, 2022; Jacovi et al., 2021; Wong et al., 2024). Further distinctions are made between cognitive, affective, and collective forms of trust, emphasising that trust is not only rational but also emotional and socially constructed (Forsyth et al., 2011; Johnson & Grayson, 2005). Studies also discuss trust development, showing how initial perceptions, system transparency, past experiences, and institutional reputation shape users' willingness to trust AI systems (Lewicki et al., 2006; Lewicki & Wiethoff, 2000). Several debates persist within the literature. One concerns the role of explainability: while many argue that systems must be transparent to earn trust (Davis et al., 2020; Thalpage, 2023; Von Eschenbach, 2021), others suggest that users often trust systems they cannot fully understand, provided outcomes appear fair or beneficial (Hieronymi, 2008; Miller, 2022; Pieters, 2011). Another debate centres on whether trust should be the goal of AI design (Jacovi et al., 2021), or whether it should emerge naturally from ethical, reliable, and user-centred systems (Ferrario et al., 2020).

These discussions reveal a gap between abstract principles and practical design strategies, particularly in complex institutional contexts like the public sector. Despite growing interest in AI and trust, there is a lack of practical research on how public services can design for trust. Much of the existing literature focuses on evaluating trust post-release or identifying the conditions under which trust fails. Less is known about how trust can be proactively fostered through design methods that address not only the technology itself but also the social systems and institutional frameworks surrounding it. This thesis seeks to fill that gap by exploring how systemic design approaches can support the development of trust in AI-enabled public services.

The study is situated in Denmark, where high public trust creates both an opportunity and a responsibility: institutions must ensure that new technologies uphold this trust, especially when adopting opaque or autonomous systems like AI (An AI Nation: Harnessing the Opportunity of Artificial Intelligence in Denmark, n.d.; Denmark Tops Europe in AI Adoption, 2025; OECD Survey on Drivers of Trust in Public Institutions 2024 Results - Country Notes, 2024; Jørgensen, 2023; Nielsen & Lindvall, 2021). Designing for

trust in this context requires sensitivity to both technical and cultural factors, and an approach that can bridge the gap between abstract principles and concrete interventions. The project is also personally motivated. Coming from India, a country which has very low trust in institutions (Kumar et al., 2021; Malik, 1979), I am particularly interested in how trust is earned and maintained in socio-technical systems. This cross-cultural perspective allows me to approach the Danish context with both appreciation and critical curiosity.

This thesis addresses the following research question:

"How might we design AI services in the public sector that ***foster trust*** *and* ***align with the needs of the end user?"***

Systemic design is especially well-suited for addressing this challenge (Jones, 2014). As a method that combines systems thinking with design practice, it enables the mapping of complex interdependencies, identification of leverage points, and facilitation of multi-stakeholder collaboration. These capacities are essential when dealing with trust, which can be considered a systemic quality that emerges from interactions between people, institutions, and technologies. By investigating this question, the study aims to contribute actionable insights for public institutions seeking to adopt AI technology into their systems responsibly, as well as to the broader academic discourse on trust, technology, and design. Ultimately, the project seeks to show how trust can be designed for, and not merely hoped for, within the development of AI-enabled public services.

1.2 Learning Goals

1.2.1 Master's Learning Goals

The master's thesis demonstrates the student's acquired competencies, skills and knowledge essential for the profession of service design (Master's Thesis (2023/2024), n.d.).

The key competencies of the student include their ability to design and develop work in complex and unpredictable situations which require innovative solutions. This includes independently initiating and implementing discipline-specific and interdisciplinary co-operation, assuming professional responsibility, and conducting technological development and research using the appropriate scientific methods to solve the cases.

The student will acquire the necessary skills to work independently, identify major problem areas, and adequately address problems and opportunities. They can analyse, design, and represent innovative solutions, as well as evaluate and address major organisational and business issues in the design of a product-service system.

Lastly, the student is expected to gain knowledge of appropriate methodological approaches for specific study areas, design theories and methods for advanced and complex product-service systems, and the relevant literature in the Service Design field (Master's Thesis (2023/2024), n.d.).

1.2.2 Personal Learning Goals

In addition to the academic objectives defined by the Master's programme, I set the following personal learning goals for this thesis:

• To explore how trust, which is a complex, abstract social value, can be supported through design in public sector contexts.

• To explore systemic design methods and thinking, and deepen my understanding of designing within complex institutional environments.

• To independently lead and manage a long-term service design project in its entirety.

• To strengthen my skills in interdisciplinary collaboration, particularly when working across professional boundaries and in a foreign language (Danish).

• To improve my confidence and competence in conducting user interviews and facilitating co-creative workshops.

• To reflect on and expand my identity as an interdisciplinary designer, drawing on my background in architecture, fashion, and service design.

2

RELATED RESEARCH

2.1 An Overview of AI

2.1.1 What is AI?

Artificial Intelligence (AI) refers to computer systems designed to perform tasks that typically require human intelligence, such as speech recognition, language understanding, learning from data, and problem-solving (Russell & Norvig, 2016). AI can be classified by capability into the following three categories:

Artificial Super Intelligence (ASI)

ASI is a hypothetical form of AI that surpasses human intelligence in all respects. It is the subject of ongoing ethical and philosophical debate (Bostrom, 2016).

Artificial General Intelligence (AGI)

AGI would match human intelligence across a wide range of tasks. It remains a theoretical idea with no working examples to date (Goertzel & Pennachin, 2007).

Artificial Narrow Intelligence (AGI)

This type of AI is designed for a specific task. Most AI tools today, such as voice assistants or search engines, fall into this group. They perform well in one area but cannot transfer knowledge to others (Kaplan & Haenlein, 2019).

Within ANI, there are several key subfields:

Rule-Based Systems

These follow strict "if-then" rules set by experts. They are predictable but limited in handling new or unclear inputs (Giarratano & Riley, 2006).

Machine Learning (ML)

ML allows systems to learn from data and improve over time without being reprogrammed. It is used in areas such as fraud detection and recommendation systems (Mitchell, 1997).

Deep Learning

A type of ML, deep learning uses multi-layered neural networks to handle complex tasks like image or voice recognition. It mimics how the human brain processes information (LeCun et al., 2015).

Natural Language Processing (NLP)

NLP allows machines to understand and generate human language. It supports chat-

bots, translation tools, and voice assistants (Jurafsky & Martin, 2000).

Large Language Models (LLMs)

These are deep learning models trained on large datasets of text. They can generate human-like language and answer questions in natural dialogue. However, they do not truly understand content and may produce inaccurate or biased responses (Bommasani et al., 2021).

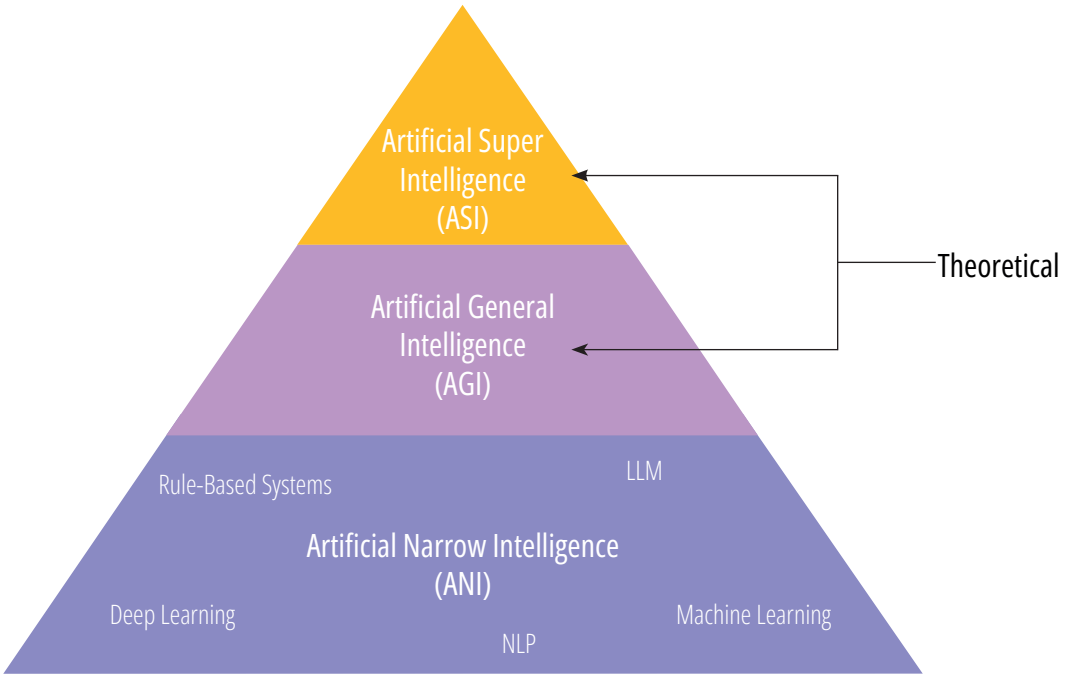


Fig. 01 Types of AI

2.1.2 AI-enabled Services

AI-enabled services use artificial intelligence technologies to deliver faster, more efficient, and personalised services. They often involve the integration of AI-tools such as chatbots, voice assistants, recommendation engines, and smart search systems, within services with the intention of enhancing user experience, automating tasks, and supporting decision-making (Wamba-Taguimdje et al., 2020).

These services are especially useful in public administration, where they can reduce manual workload, improve consistency, and make services more accessible to the public. Public sector organisations are increasingly adopting these AI technologies to improve service delivery, optimise operations, and support policy implementation (Chun, 2008; De Fine Licht & De Fine Licht, 2020; Wirtz et al., 2019). These AI service systems are used in a range of areas, including welfare case management, citizen inquiries, traffic control, tax administration, and healthcare (Eggers et al., 2017; Mergel et al., 2023; Ubaldi et al., 2019). AI-enabled services often reduce administrative burden, improve decision-making speed, and increase access to healthcare.

Common examples of AI-enabled services include:

Chatbots and Virtual Assistants

Used by governments and companies to answer routine questions or guide users to the right resources (McTear, 2021).

Smart Search Engines

Help employees or users find the most relevant information from large databases, often using natural language queries (Shroff, 2013).

AI in Customer Service

Assists agents by suggesting responses or resolving simple queries automatically (Davenport & Ronanki, 2018).

Voice Interfaces

Enable users to interact with services using spoken language, such as with voicebots in call centres (Murad & Munteanu, 2020).

However, introducing AI-enabled services within public organisations differs significantly from private sector use. Public services must uphold principles of equity, transparency, and accountability. AI-enabled services that affect citizen rights or entitlements must be not only efficient, but also explainable, contestable, and fair (Bostrom & Yudkowsky, 2011). Public services often comprise high-stakes environments, where errors can lead to severe consequences for individuals. Examples include AI tools used to assess eligibility for social benefits or prioritise medical treatment (Booth & correspondent, 2024; Booth & editor, 2024). This heightens the need for public trust and careful system design. Moreover, government frameworks for ethical oversight and risk management are still evolving (Jobin et al., 2019; Jozak, 2025).

2.1.3 Challenges in public AI-enabled Services

While AI-enabled services offer many benefits - such as efficiency, accuracy, and automation - they also come with several challenges. These are especially significant in the public sector, where services must be transparent, fair, and accountable.

Transparency and Explainability

Many AI systems, especially those based on machine learning, work as “black boxes”. This means it can be difficult to understand how they arrive at a decision or recommendation (Burrell, 2016). In the public sector, this poses a serious problem. Public institutions must be able to explain and justify their actions to citizens. If an AI system denies a service or makes an error, both the employees and the citizens may not understand why it happened (Wirtz et al., 2019).

Data Quality and Bias

AI systems learn from data. If the data is biased, incomplete, or outdated, the AI may

make unfair or incorrect decisions (Mehrabi et al., 2022). In the public sector, poor data quality can lead to inequality in access to services. For example, if a system is trained mainly on data from one group, it may not work well for others. This could reinforce existing social biases or discrimination (Eubanks, 2019).

Ethical and Legal Concerns

AI raises several ethical and legal questions. Who is responsible if the system makes a mistake? How do we protect citizen privacy and data security? What rules should guide AI use in sensitive areas like welfare or healthcare? Public sector organisations often lack clear policies for managing these risks (Wirtz et al., 2019).

Skills and Organisational Readiness

Developing and using AI services requires technical skills, as well as new ways of thinking about service design and delivery. Many public institutions lack the required expertise or resources (Ubaldi et al., 2019). In addition, existing processes and organisational cultures may not be ready to work with AI.

Cost and Complexity

AI systems can be expensive to develop, maintain, and update. They also require constant oversight to ensure accuracy and fairness. For public institutions, this can be a barrier, especially when budgets are limited and priorities are competing (Medaglia et al., 2023).

Trust and Adoption

AI tools often face resistance from employees and citizens. If people do not trust the technology, they may avoid using it or use it incorrectly (Eggers et al., 2017). In public organisations, trust is key. Citizens must feel that AI systems are fair and reliable, and employees must trust that the tools support their work, not replace them (Margetts & Dorobantu, 2019).

If these challenges are not properly addressed, they may lead to mistrust, harm, or increased inequality. A responsible and inclusive approach is needed to ensure AI serves the public in a responsible and fair manner.

2.1.4 From Risk to Trust: Why Trust in AI Matters

The challenges surrounding AI-enabled services pose serious risks to their effective deployment. As previously mentioned, these issues are particularly acute in the public sector, where decisions must be accountable, equitable, and understandable to both citizens and employees. Trust becomes the cornerstone of successful implementation. Without it, citizens may reject the outcomes of AI-supported decisions, and frontline staff may resist using the tools provided to them (Eggers et al., 2017; Wirtz et al., 2019). Trust, therefore, is not a secondary concern or an optional feature. It is a necessary con-

dition for the responsible, ethical, and sustainable use of AI in public services (De Fine Licht & De Fine Licht, 2020; Wilson & Van Der Velden, 2022). It connects directly to the public's willingness to engage with AI and the public organisation's ability to deliver services that are both efficient and legitimate. In the following section, I explore what trust in AI entails, why it matters, and how it can be deliberately supported through design.

2.2 Trust in AI

2.2.1 What is Trust?

Webster's dictionary defines trust as the *"assured reliance on the character, ability, strength, or truth of someone or something"* and *"one in which confidence is placed"* (Definition of TRUST, 2025). Trust has been extensively studied across disciplines such as psychology, sociology, and organisational theory. At its core, trust is the willingness to accept vulnerability based on positive expectations of the other party's behaviour or intentions (Rousseau et al., 1998). In the context of technology, this often involves confidence that a system will function reliably, ethically, and transparently even in uncertain or high-stakes situations (Lankton et al., 2015; X. Li et al., 2008). Taddeo (2009) goes further to define this as e-trust. He describes this trust as that which is formed in environments mediated by digital devices, where moral and societal expectations are differently perceived, and where there is no direct and physical contact. Based on his study, it is clear that trust is not limited to the technology itself. It encompasses:

- Trust in the **system's outputs** (are decisions accurate, fair, explainable?),
- Trust in the **institutions behind the technology** (are they accountable and transparent?), and
- Trust in the **process** (was the AI designed and deployed responsibly?).

2.2.2 Types of Trust

Scholars have identified multiple forms of trust relevant to AI-enabled public services:

Interpersonal Trust

Interpersonal trust is directed towards other individuals, such as frontline staff, experts, or decision-makers. This type of trust is often built through personal interaction and perceptions of integrity, competence, and benevolence (Jacovi et al., 2021).

Technological Trust or E-trust

E-trust is directed towards digital tools and systems, like AI itself. It relates to the perceived reliability, functionality, and transparency of the system. Users must feel that the technology performs as expected and is free from hidden risks (Taddeo, 2009).

Institutional Trust

Institutional trust refers to confidence in the frameworks, rules, and organisations responsible for a system's design and oversight (Sønderskov & Dinesen, 2016). It is par-

ticularly important in the public sector, where legitimacy and accountability are central (Bedué & Fritzsche, 2022).

Cognitive Trust

Cognitive trust is based on rational assessment of competence and reliability (Johnson & Grayson, 2005). It is often grounded in perceptions of the system's performance, accuracy, and technical robustness (Glikson & Woolley, 2020).

Affective Trust

Affective trust arises from emotional bonds, empathy, or perceived goodwill (Johnson & Grayson, 2005). While less discussed in technical literature, it plays a role in how users interpret automated decisions, especially when services involve care or vulnerability (Kyung & Kwon, 2022; M. Ryan, 2020).

Collective Trust

Collective trust reflects shared beliefs within a group or society about the trustworthiness of institutions, technologies, or decisions. It is shaped by public discourse, social norms, and historical experience. Collective trust is crucial in public administration, where services must be seen as trustworthy not just by individuals, but by communities at large (Kramer, 2010; Kramer et al., 1996).

The types of trust mentioned above are just a few of the dimensions that influence our trust in AI. Is it essential to understand that this trust is not a static attribute, but a dynamic relationship shaped by design, communication, user experience, and organisational and social context.

Fig. 02 Types of trust (clockwise, starting from top left): interpersonal, e-trust, institutional, cognitive, collective, affective.

2.2.3 Development of Trust

Trust is not static; it evolves through use, interaction, and context. In the context of AI in the public sector, trust must be deliberately cultivated. One of the most widely referenced frameworks is Mayer et al.'s (1995) integrative model of organisational trust, which can be applied to the public sector. According to this model, trust is based on the perceived ability, benevolence, and integrity of the party being trusted.

• **Ability** refers to the technical competence of the AI system or the organisation behind it. Users need to believe the system is capable of performing its intended function effectively.

• **Benevolence** reflects the belief that the system or its developers act in the user's interest. This is especially critical in public services, where citizens must feel the system is designed to serve the common good.

• **Integrity** involves the perception that the system and its operators adhere to principles that are acceptable to the trustor, such as fairness, transparency, and ethical standards.

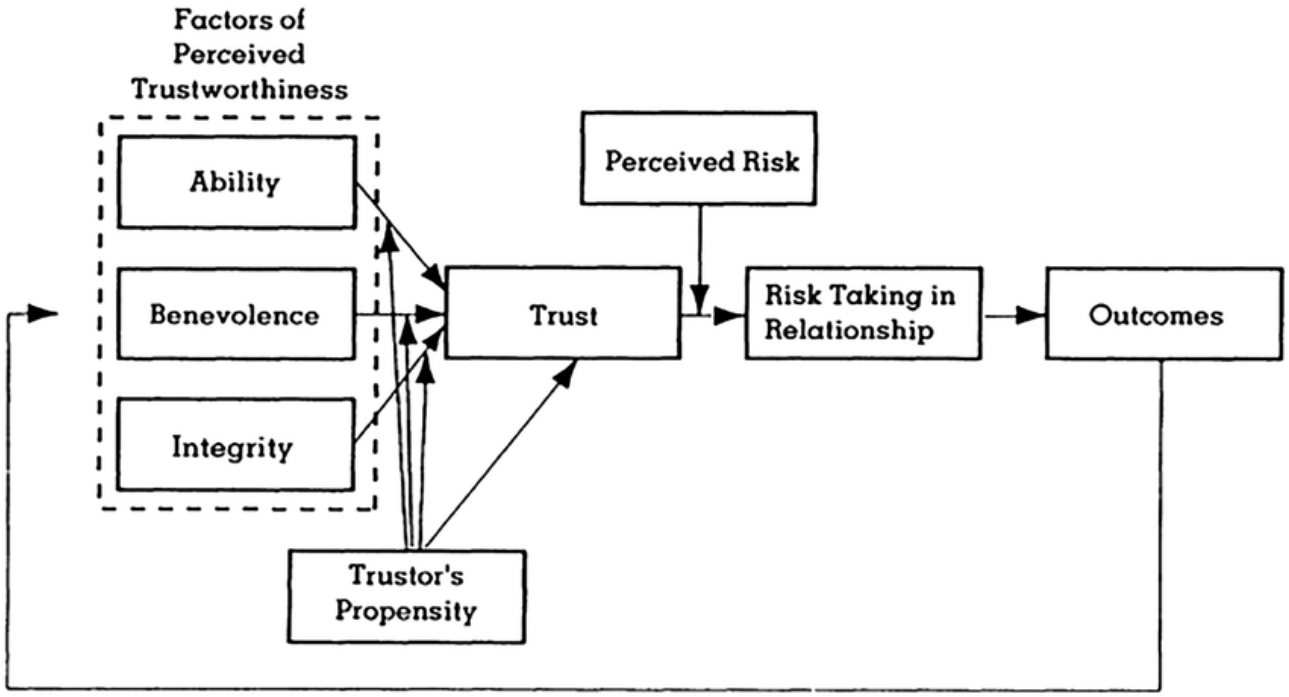


Fig. 03 Mayer et al.'s Model of Trust
Source: Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. The Academy of Management Review, 20(3), 709. https://doi.org/10.2307/258792

Beyond this model, several conditions support trust development in AI-enabled public services:

Transparency and Explainability

When users can understand the rationale behind system decisions, trust increases (Guo, 2020; Schmidt et al., 2020).

Reliability and Predictability

Repeated, consistent performance reinforces trust. Unexplained errors quickly erode it (M. Ryan, 2020).

Fairness and Non-Discrimination

Users must believe that decisions are impartial and do not reinforce bias or inequality (Lee & Rich, 2021).

Accountability and Oversight

Trust is reinforced when mechanisms exist to contest decisions, assign responsibility, and intervene when needed (Mardiani et al., 2023).

Engagement and Inclusion

Involving users and stakeholders in AI design and feedback loops creates a shared sense of ownership and legitimacy (Delgado et al., 2023).

Trust development is comprehensive, but fragile. It builds slowly through positive interactions but can be lost quickly when systems behave unpredictably, opaquely, or unfairly, especially in high-stakes public contexts.

2.2.4 Denmark's Relationship with Trust and Technology

Denmark is often described as a high-trust society, with strong institutional trust in government, public services, and the rule of law (Larsen, 2013; Nielsen & Lindvall, 2021; OECD Survey on Drivers of Trust in Public Institutions 2024 Results - Country Notes, 2024; Sønderskov & Dinesen, 2014). This cultural foundation of trust extends to the adoption of new technologies in the public sector. The Danish public tends to have confidence in authorities to act in the public interest, which facilitates acceptance of government-led digital initiatives (Denmark Tops Europe in AI Adoption, 2025; Jørgensen, 2023). This has contributed to Denmark's leadership in digital governance and citizen-facing e-services. The introduction of platforms such as Borger.dk or MitID has seen high adoption and relatively low resistance, partly due to the belief that systems are built transparently and with user welfare in mind (MitID Is Denmark's Digital ID - MitID, n.d.). However, trust cannot be assumed to transfer automatically to more complex technologies like AI (Lorenzen, 2024). While there is baseline trust in institutions, emerging AI systems raise new questions around transparency, fairness, and oversight. Studies show that Danish citizens support digital innovation but expect public institutions to maintain strong ethical standards and provide clear accountability (van Kersbergen & Tinggaard Svendsen, 2024). In this context, trust in technology is shaped by three interrelated dimensions: cultural norms of interpersonal and institutional trust, the state's role as a trusted digital innovator, and the public's expectations for fairness, explainability, and control. As Denmark continues to develop AI-enabled services, maintaining this delicate balance

between innovation and public trust will be essential.

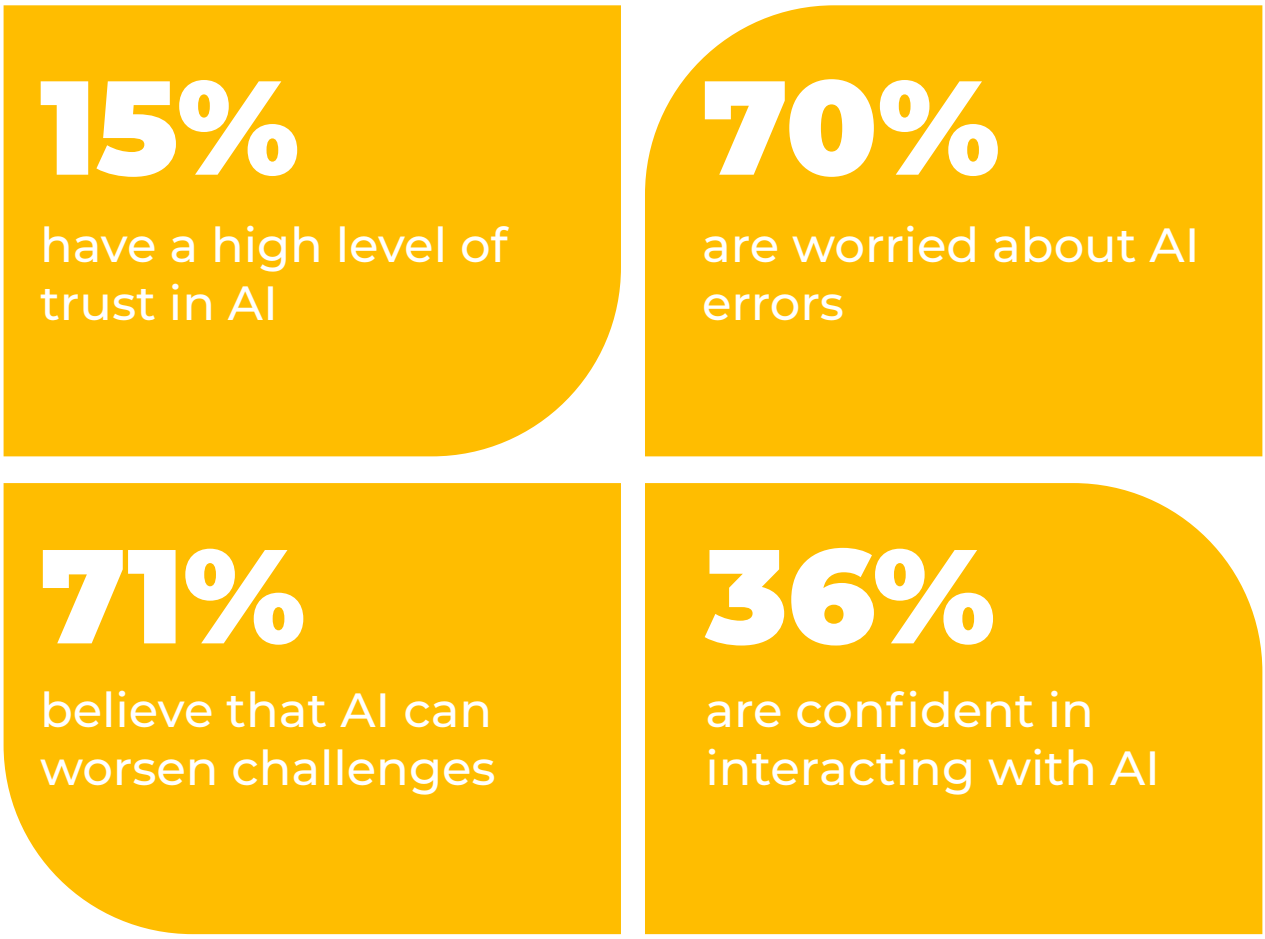


Fig. 04 Results from the survey conducted by Lorenzen, M. S. (2024). Digital Democracy Centre Undersøger: Danskernes forståelse af, holdninger til og brug af (generativ) kunstig intelligens. University of Southern Denmark. https://doi.org/10.21996/GPMZ-S343

2.2.5 Diverging Perspectives on Trust in AI

While there is broad consensus that trust is essential to the successful implementation of AI in public services, several key debates persist in the literature regarding how trust should be understood and operationalised.

A central discussion concerns the role of explainability in cultivating trust. On one hand, many scholars argue that users must be able to understand how AI systems work (or at least grasp the logic behind decisions) in order to trust them (Guo, 2020; Miller, 2022; Shin, 2021; Von Eschenbach, 2021). Explainable AI (XAI) is thus promoted as a technical solution to the trust problem, especially in high-stakes domains such as healthcare or public administration. However, others challenge the assumption that transparency necessarily leads to trust (Ananny & Crawford, 2018; Kroll et al., n.d.; Weller, 2019) . They point to the fact that users often trust complex systems, including human experts and institutions, without fully understanding their inner workings, provided the

outcomes are consistent, fair, and beneficial. This suggests that perceived competence and fairness may sometimes outweigh the need for detailed technical transparency. A second debate addresses whether trust should be explicitly designed for, or whether it should emerge as a natural consequence of ethical, accountable, and user-centred systems. Some authors warn against designing with trust as a "manipulative goal," arguing that efforts to build trust must be earned through performance, governance, and accountability rather than persuasive design (Avin et al., 2021; Liao & Sundar, 2022). Others, however, advocate for trust to be treated as a designable quality, calling for new methods and tools that integrate trustworthiness into the service (Hoff & Bashir, 2015; Riegelsberger et al., 2005). Another debate revolves around whether it is even appropriate to talk about trusting AI systems in the same way we trust humans or institutions. Some argue that trust is a fundamentally human phenomenon, based on concepts like intention, empathy, or moral responsibility, which machines do not possess (Coeckelbergh, 2012; Hancock et al., 2011). From this perspective, users may only ever rely on or depend on machines, not truly trust them. Others propose that as AI systems increasingly take on decision-making roles, we must adapt our understanding of trust to fit human-machine relations (M. Ryan, 2020; Taddeo, 2009). Finally, some perspectives view distrust as inherently negative, a barrier to overcome in AI adoption (Siau & Wang, 2018). Others argue that healthy scepticism is essential to democratic governance, especially in public sector settings where power and accountability are at stake (Ananny & Crawford, 2018; Levi & Stoker, 2000). From this view, a certain degree of scrutiny and critical engagement is not only expected but necessary and efforts to "design for trust" should not aim to eliminate doubt, but to support informed trust through checks, balances, and public oversight mechanisms.

These debates point to a broader gap in the literature: while trust is frequently discussed in theoretical and ethical terms, there is limited practical guidance on how to support trust through the design of AI-enabled services in the public sector. This gap highlights the need for interdisciplinary, context-sensitive approaches that move beyond abstract principles and engage with the real-world complexities of public sector innovation.

3

METHODOLOGICAL APPROACH

To address the complex, systemic nature of AI adoption and trust in public services, I combined a Systemic Design (Jones, 2014) approach with the Double Diamond Framework (Ball, 2004). This hybrid methodology supported both user-centred exploration and broader system-level understanding.

3.1 The Double Diamond Framework

The *Double Diamond Framework (DDF)* is a design framework developed by the Design Council as a visual representation of the design and innovation process.

It is a simple framework comprising two diamonds – the first representing the problem space (solve the right problem) and the second the solution space (solve the problem right). Each diamond is divided into two phases, yielding a total of four phases: Discover, Define, Develop, and Deliver. The framework incorporates convergent and divergent thinking within each phase. Divergent phases, such as Discover and Develop, focus on generating a wide range of ideas, while convergent phases, like Define and Deliver, narrow down these ideas to find the most suitable solutions.

While the DDF presents certain limitations, mainly that it is a linear model (Lawson, 2010), it remains useful as a high-level structure. In this project, the DDF serves as a guiding scaffold to manage the process and communicate its progression, while acknowledging that real-world design practice is often iterative and non-linear.

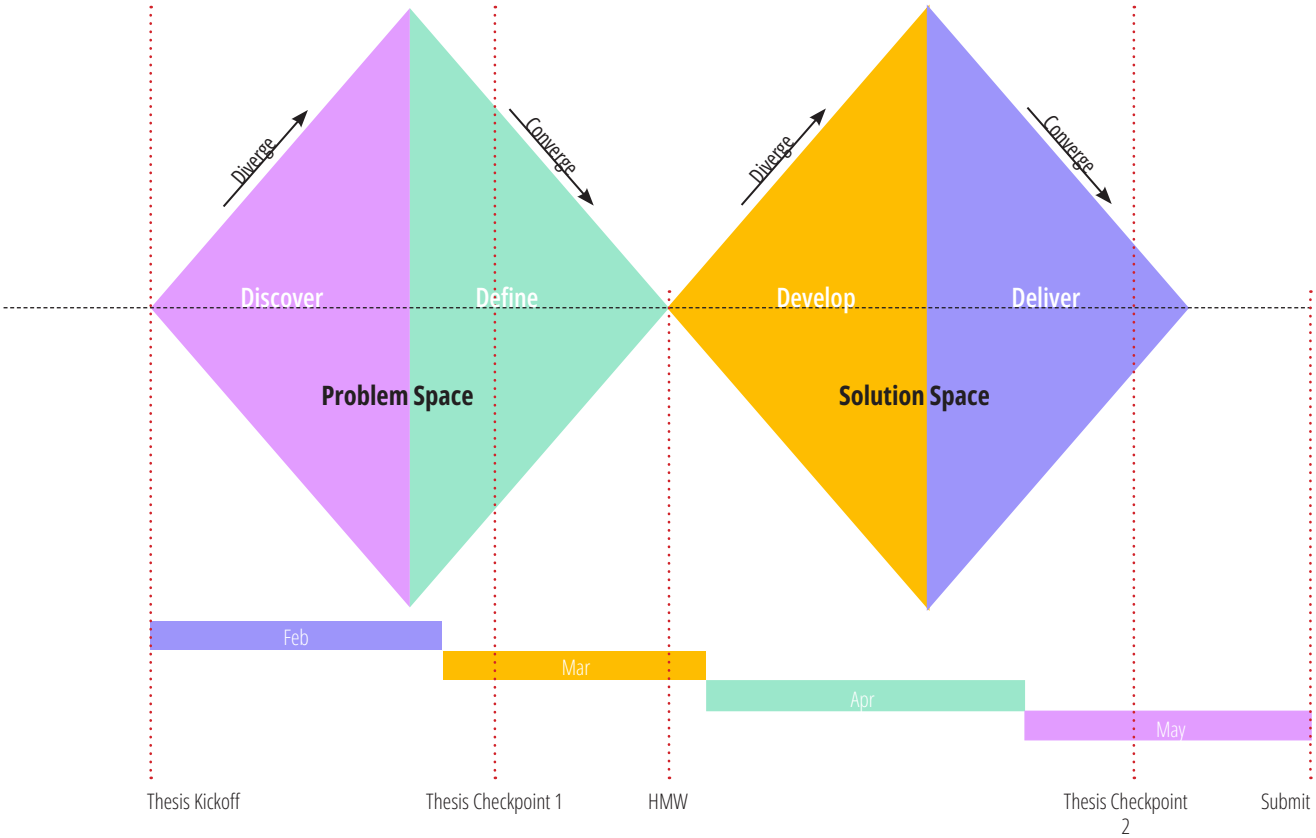


Fig. 05 The Double Diamond Framework adapted for this project

3.2 Systemic Design

To navigate the complex context of AI service adoption in a large public organisation, I will apply a Systemic Design approach to the DDF. Systemic Design is a combined approach that combines the human-centred focus of Design Thinking (Brown, 2008) with the holistic perspective of Systems Thinking (Richmond, 1993). By doing so, a Systemic Design approach can enable design solutions that are both user-friendly and contextually stable.

The methodology is nonlinear and iterative in its application, encompassing six activities: inquiring, framing, formulating, generating, reflecting, and facilitating (Battistoni et al., 2019). This allows for a dynamic and iterative process, where both problem framing and solution development evolve together. This approach is particularly relevant for this project as it allows for the inclusion of multiple stakeholder perspectives, including service creators and front end employees, and supports interdisciplinary collaboration. Indeed Systemic Design deliberately pursues interdisciplinary collaboration by bringing together different stakeholder types, which can facilitate innovative solutions (Battistoni et al., 2019). A Systemic Design approach facilitates a collective understanding of a situation that is too complex to be fully understood by a single perspective (A. Ryan, 2014). Thus, it will help enable a more holistic framing of trust - not as a standalone issue, but one influenced by workflows, communication, and broader organisational culture.

4

DESIGN CASE

4.1 About the Organisation

The foundation for this thesis was laid during the ninth semester of the Service Systems Design master's programme, when I carried out my internship in the department of Citizen Service Development within the Copenhagen Municipality and continued into the tenth semester as my thesis project. The following sections outline the organisational structure and the roles of the relevant departments.

4.1.1 Copenhagen Municipality

Copenhagen Municipality or Københavns Kommune is Denmark's most populous municipality, home to 659,350 residents as of January 1, 2024 (Befolkningen 1. Januar - Statistikbanken, n.d.). It is divided into 10 districts and governed by seven administrative branches, each responsible for a specific domain. These include: (1) The Economic Administration, (2) The Culture and Leisure Administration, (3) The Children and Youth Administration, (4) The Health and Care Administration, (5) The Employment and Integration Administration, (6) The Technical and Environmental Administration, and (7) The Social Administration (Employee.Kk.Dk | Employee in Copenhagen Municipality, n.d.).

During my internship, I was part of the Culture and Leisure Administration, which comprises an executive board and three centres: (1) Citizen Service and Authorities, (2) Cultural and Leisure Activities, (3) Finance, Digitization and Organization (The Organization of the Culture and Leisure Administration | Copenhagen Municipality, n.d.). The 'Citizen Service and Authority' centre is further divided into two areas: Authority and Internationalization and Citizen Service. Finally, the Citizen Service area consists of the following areas: "The Citizen Helpline" (phone service centres for the citizens), "Citizen Service in the City" (seven physical Citizen Service Centres across Copenhagen), and the "*Citizen Service Development (CSD)*" department (responsible for developing services for citizens).

The first two deliver services to citizens directly, while CSD focuses on service improvement and innovation. These services include passport issuance, MitID and digital services, driving licences, address registration, pensions, marriage services, name registration, and personal data services (Københavns Borgerservice | Københavns Kommune, n.d.).

This thesis was carried out in collaboration with the department of CSD.

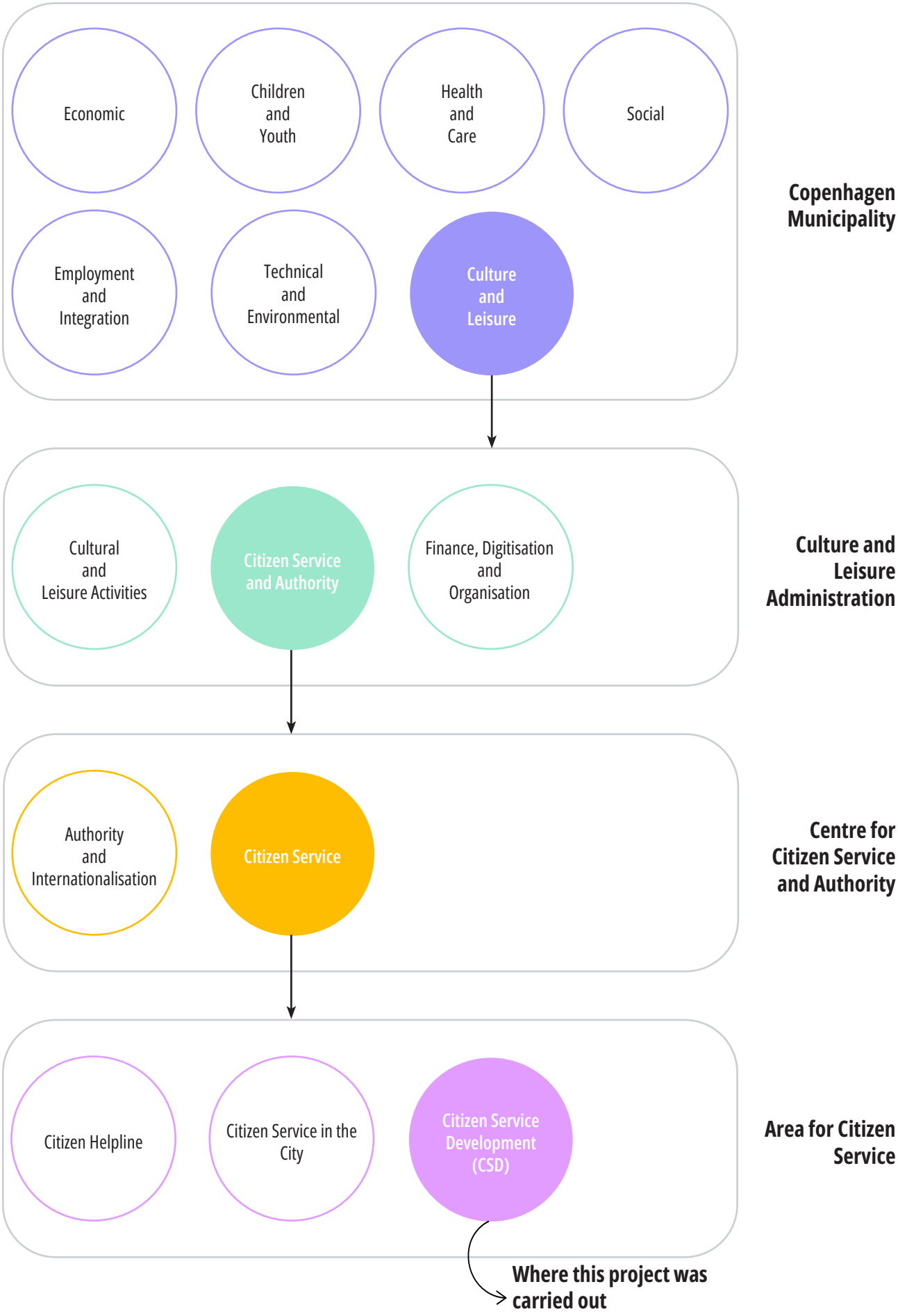


Fig. 06 Copenhagen Municipality Organisational Structure

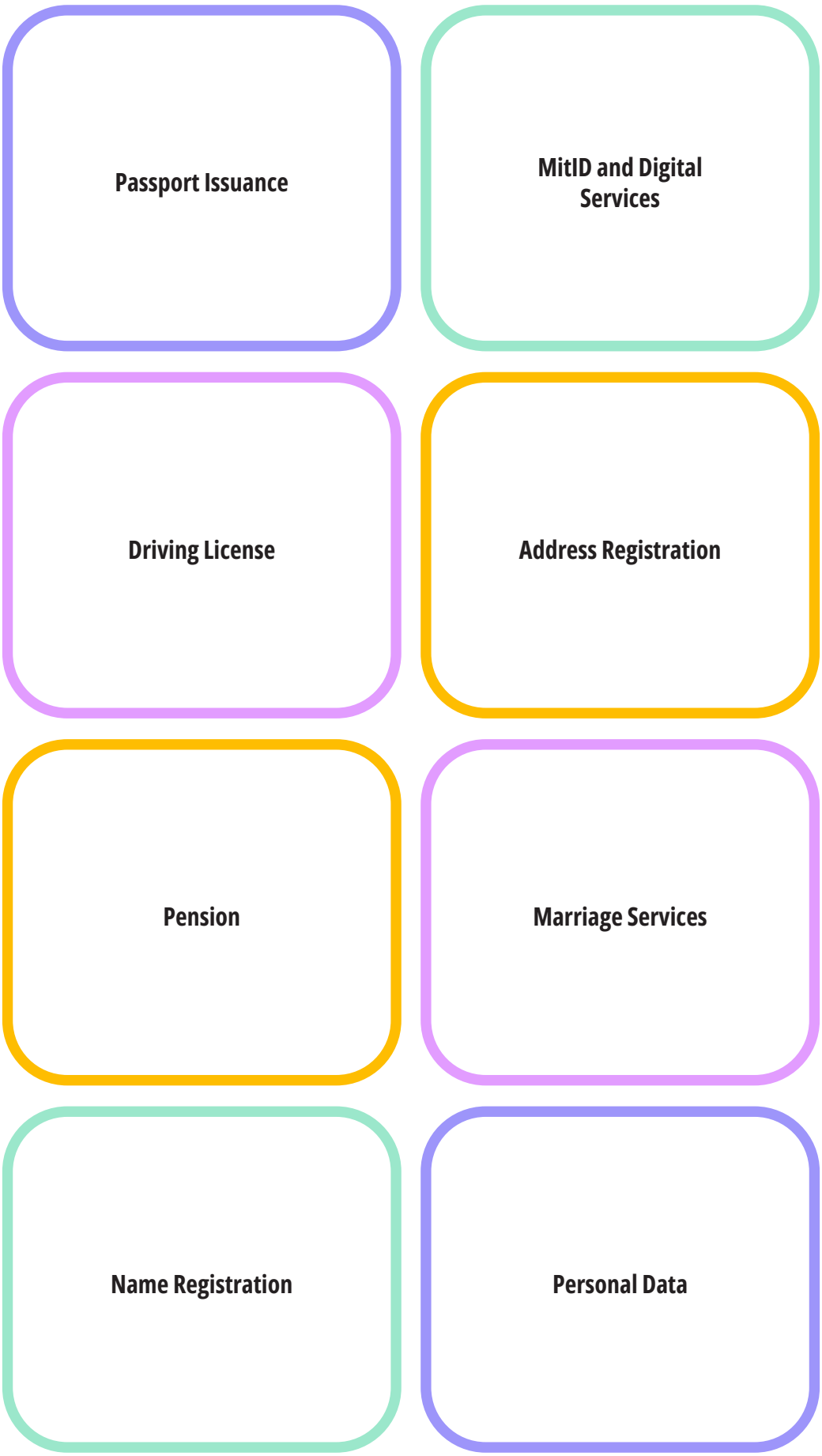


Fig. 07 Services offered by the Area for Citizen Service

4.1.2 Department of Citizen Service Development

Citizen Service Development (CSD) is the department responsible for innovating and improving public services offered to the citizens of Copenhagen. This department often collaborates closely with the Citizen Helpline and Citizen Service in the City to engage front-end employees and citizens in the optimisation process. CSD comprises 23 employees with diverse roles, including Project Managers, Data Analysts, Development Consultants, Digital Development Consultants, Communication Experts, Web Editors, IT Specialists, and Service Designers. Their work spans four areas: operations, development, secretariat functions, and consultancy services for other departments. Some tasks focus on improving current workflows (e.g., the booking system, service layout), while others aim to align services across different municipal administrations.

One of the key ongoing projects at CSD is the Digitalisation Strategy 2027, which aims to integrate AI-enabled solutions into existing services (Digitaliseringsstrategi 2024-2027 – Et bedre københavnerliv, n.d.). By 2027, AI will play a central role in enhancing self-service options for citizens, reducing reliance on manual processes, and freeing up employee resources for more complex tasks. CSD is responsible for developing and testing several AI-enabled tools under this strategy, in collaboration with other departments. Three AI-enabled solutions under development are:

- **AI-Chatbot:** a chatbot integrated into the municipality's website to assist citizens with inquiries, providing instant responses and guiding them to relevant services.
- **AI-Voicebot:** a voicebot for the Citizen Helpline to handle routine calls, improving response times and freeing up employees for more complex cases.
- **AI-Searchbot:** a tool designed to replace the current internal search engine used by front-end employees to find guidance on handling citizen queries when procedures are unclear. The searchbot aims to streamline workflows, improve response accuracy, and enhance service efficiency.

While the tools differ in use and function, they are all part of a broader shift towards AI-enabled public services. These tools rely on artificial intelligence to process information and generate responses in real time.

4.1.3 Focus Area: AI-Searchbot

This thesis focuses on the AI-searchbot: a tool which is developed by CSD in collaboration with Koncern IT (KIT), the department responsible for delivering IT services across the municipality. The development began in 2024 and the project aims to replace the current internal search engine used by front-end employees to look up procedures when handling citizen queries. The existing system functions like a basic keyword search and often returns results written in highly technical language, making it difficult to use efficiently.

The AI-Searchbot is powered by a large language model (LLM), a type of AI that is trained to understand and generate human-like text. This capability stems from advancements in neural network architectures, particularly the Transformer model (Vaswani et al., 2017), which enables more coherent and context-aware dialogue. Unlike self-learning AI-systems, however, this tool does not adapt or update its knowledge based on user interaction. It retrieves and reformulates information from a fixed internal database, making it a static but more user-friendly alternative to the previous system. The AI-searchbot is currently being rolled out to employees in a pilot phase. It is designed to streamline internal workflows, improve the consistency of responses, and support service quality.

The tool is being rolled out to front-end employees across six different teams: The Citizen Helpline, Citizen Service in the City, Jobcenter Copenhagen, the Technical and Environmental Administration, Parking, and Copenhagen Business House. End users were involved to a limited extent during development, mainly through testing, validation, and participatory UI design. At the time of this thesis, the AI-searchbot was in the early- to mid-phases of being rolled out to the end-user.

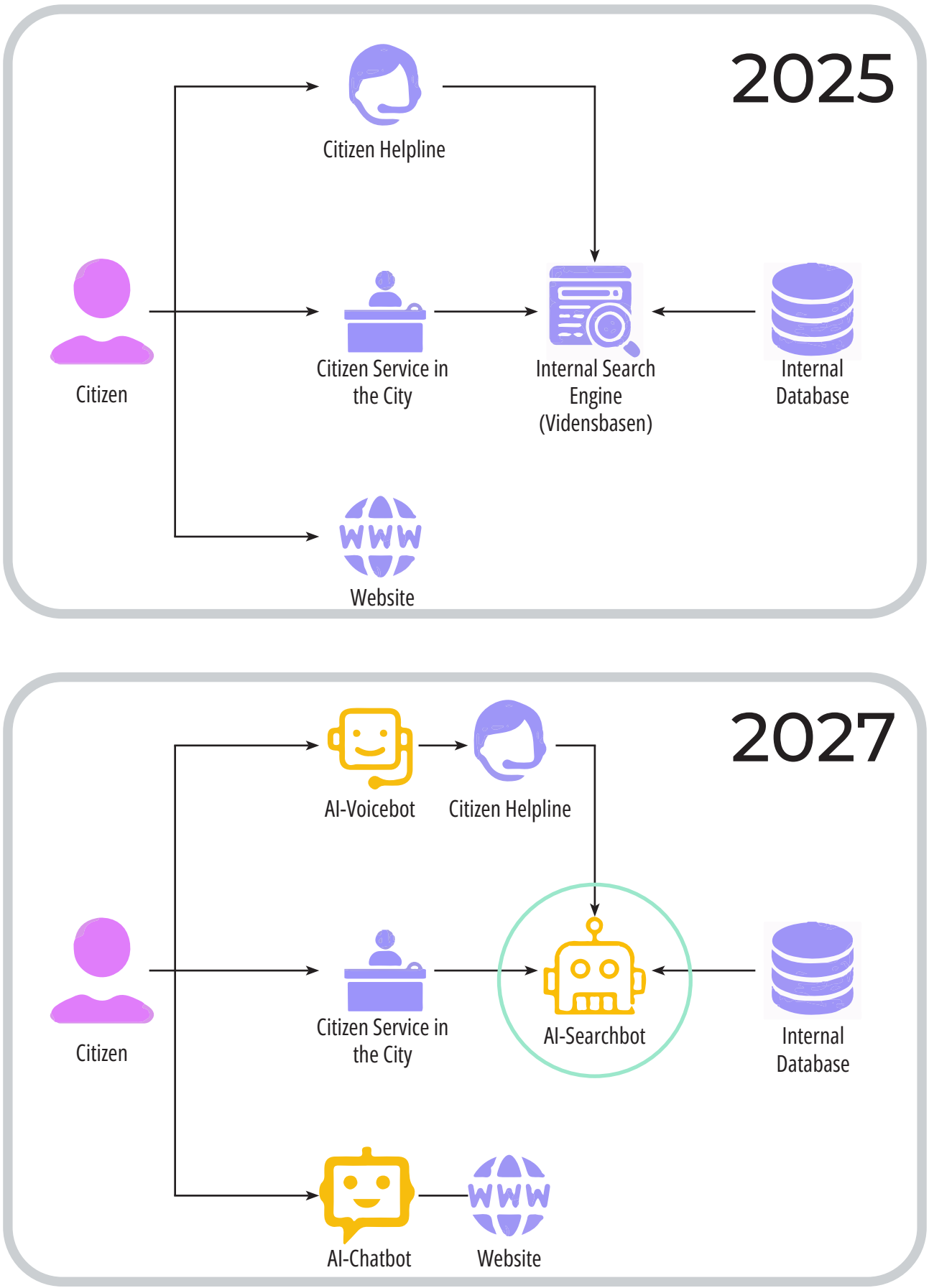


Fig. 08 Citizens interacting with the various touchpoints of Citizen Service today and in 2027

4.2 Brief

The brief for this project was developed collaboratively between CSD and myself, aligning with the department's needs, the university programme's curriculum, and my interests.

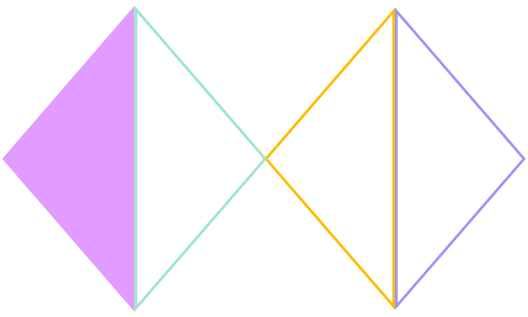
The brief focused on understanding the challenges service creators at CSD faced during the rollout and adoption of the AI-searchbot, with the primary aim of exploring how CSD could support its adoption without undermining front-end employee trust.

The goal of this work is to provide actionable insights and recommendations that will help CSD successfully integrate the AI-searchbot into daily operations while building employee trust and encouraging adoption. The findings may also support the rollout and adoption of other AI-enabled services within the department and municipality.

5

DESIGN PROCESS

5.1 DISCOVER



About This Phase

In this phase, I explored the current landscape of service delivery and AI-enabled service development in the Citizen Service area of Copenhagen Municipality. I conducted observations and interviews with key stakeholders, including service creators at CSD and front-end employees from Citizen Service Centres and the Citizen Helpline. To gain a broader perspective, I also spoke with individuals from other organisations who had experience designing for trust in AI-enabled services. This phase helped me identify the different actors involved in the service ecosystem, which I visualised through a stakeholder map. While gathering these diverse perspectives, I gradually narrowed my focus to the service creators at CSD, who became the primary target group for this project.

5.1.1 Understanding the End-User of the Service

5.1.1.1 Observations

To explore how front-end employees might adopt and trust the AI-searchbot, I examined their behaviours, work patterns, and problem-solving strategies across three distinct service areas. Drawing on qualitative fieldwork from my roles as an intern and student helper at CSD, I observed how staff in Copenhagen Municipality's Citizen Service Centres interact with citizens and deliver services under varying conditions. The aim was to identify behavioural patterns, particularly contrasting routine and complex tasks, and human- versus technology-driven service delivery.

The observations focused on three core services:

- **MitID** (the national digital ID) (About MitID - MitID, n.d.)
- **Kørekort** (driving licence) (Tag et kørekort | Københavns Kommune, n.d.)
- **Digital Fuldmagt and Digital Post** (services enabling access to digital platforms on behalf of others) (Digital Fuldmagt | Københavns Kommune, n.d.; Digital Post | Københavns Kommune, n.d.).

I conducted in-situ observations at two citizen service centres - Vesterport and Sundby, where I observed both front-office and back-office employees over a period of approximately two weeks for each service type. In addition to live observations, I reviewed recorded telephone conversations between Citizen Helpline employees and citizens to complement and triangulate the findings.

MitID emerged as a routine and highly standardised service. Citizens typically contacted the municipality to activate or reactivate their MitID, often due to lost devices or account security issues. I observed 15 MitID service operations involving seven different employees, and reviewed approximately 30 recorded citizen interactions via the helpline. In nearly all cases, employees carried out the service confidently and independently, relying on personal experience rather than internal systems or colleagues. The tasks were procedural, with fixed solutions and minimal variation. By contrast, the Kørekort service proved to be significantly more complex. While some citizens had routine queries (such as applying for a first-time driving licence), others presented cases involving reacquisition or the exchange of non-EU licences, which were processed only at the Vesterport centre. These complex services were often handled by a small group of specialists who possessed comprehensive knowledge of the domain. During a three-day design sprint focused on this team, we observed that although employees attempted to use the internal knowledge database, its technical language made it difficult to navigate. Consequently, front-line staff often turned to specialists for guidance. A supplementary two-week survey revealed that specialists were contacted by Citizen

Helpline staff approximately every 20 minutes, indicating a high level of dependency and frequent workflow interruption. The final category of observation covered Digital Fuldmagt and Digital Post, which are services that enable a trusted person to act on behalf of a citizen, particularly in situations involving illness, cognitive decline, or other vulnerabilities. These services often overlap in practice and are part of what is internally referred to as the 'Social Area' of citizen services. I observed 12 service operations across both centres and reviewed ~15 recorded helpline conversations. These cases were frequently complex and time-consuming, with service durations extending up to an hour. The greatest challenge for employees was not the technical execution, but rather understanding the citizen's situation, clarifying their needs, and identifying the appropriate service response. In this context, employees regularly turned to one another for guidance and consulted the internal database for reassurance or clarification. Here, peer support and access to shared knowledge were vital for delivering appropriate and empathetic service.

Visual documentation of front-end and back-end employee observations

5.1.1.2 Mapping Behaviours

To better understand how employees interacted with resources during service delivery, I mapped a typical user journey (Rosenbaum et al., 2017) of Citizen Service employees across key touchpoints. Drawing on Patrício's Multilevel Service Design framework (2011), the mapping considered interactions at the service, interaction, and resource levels. It revealed how front-end employees engaged with internal tools, peer support, or a combination of both to identify and respond to citizens' needs. As shown in the illustration, resource use was often prompted by the need to clarify the underlying issue, with the employees selecting human and/or technological support, based on the complexity and nature of the situation.

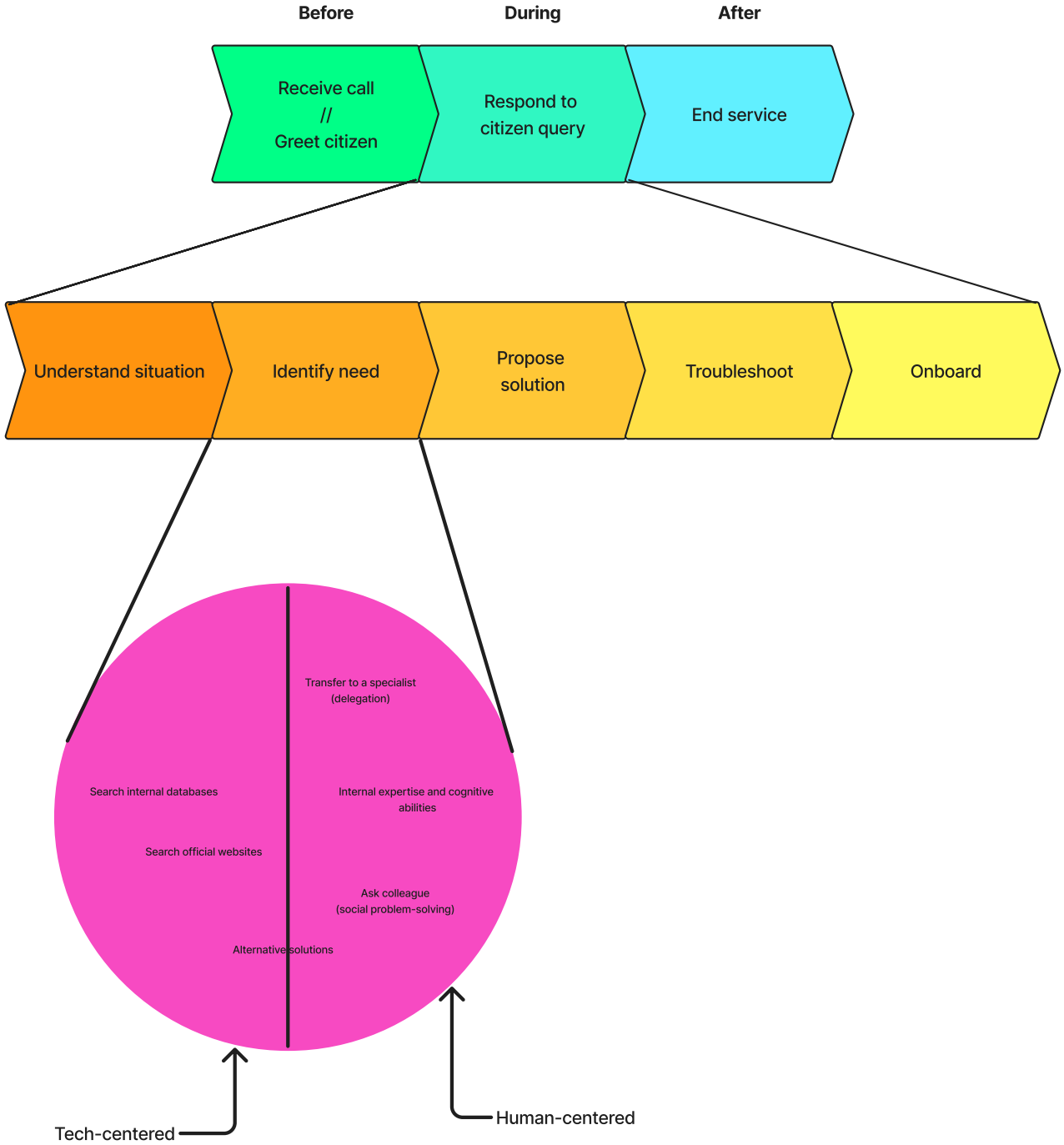


Fig. 09 User journey for front-end employees of Citizen Service

To further analyse behavioural patterns more systematically, I mapped the interactions for MitID and Digital Fuldmagt/Post, which were observed in the Citizen Service in the City Centres onto individual 2x2 matrices. Each matrix used the same axes: 'tech-driven to human-driven' (x-axis) indicating whether employees were using technological or human resources and 'routine to complex' (y-axis) indicating the level of complexity of that particular service operation. This approach allowed for comparative insight into how employees navigated each service type.

MitID cases clustered in the routine/human-driven quadrant. This revealed that employees followed predictable workflows and rarely consulted others, drawing on internalised procedural knowledge. Digital Fuldmagt and Digital Post cases were mapped and found to be largely complex, with employees using both tech- and human-driven resources to identify the underlying issues during citizen queries. The clustering revealed that employees relied on both internal systems and peer support to interpret and respond to citizens' nuanced and often sensitive needs. For the driving license service, no matrix was constructed due to the limited observational data. However, an internal survey conducted with service specialists provided insight into the demanding and the dependent nature of expert support. Unfortunately, that particular data is confidential and cannot be shared in this report.

Together, these findings suggest that the successful adoption of the AI-searchbot will depend on how well it integrates with existing cultural orientations, namely: procedural autonomy in routine tasks, informal peer reliance in complex, context-driven services, and specialist dependency where internal resources are insufficiently accessible. These orientations shape not only service delivery, but also how new tools are likely to be received, trusted, and integrated into everyday practice.

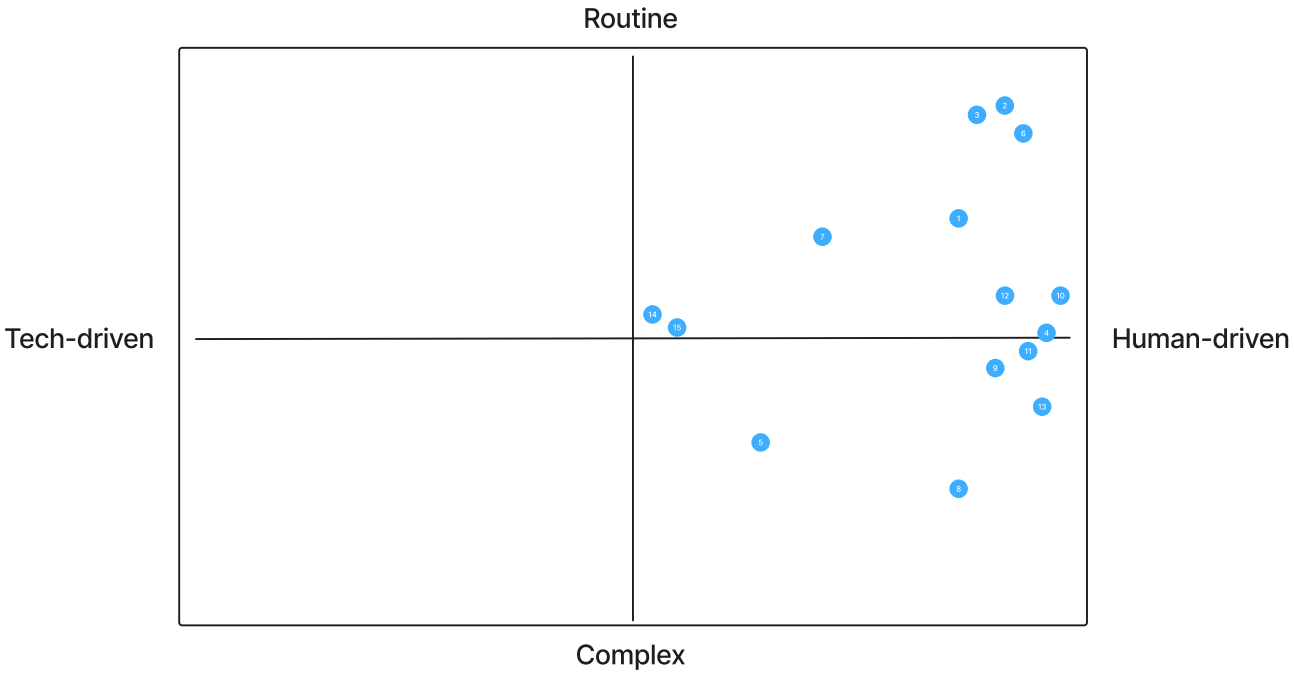


Fig. 10 Clustering of MitID Service Operations

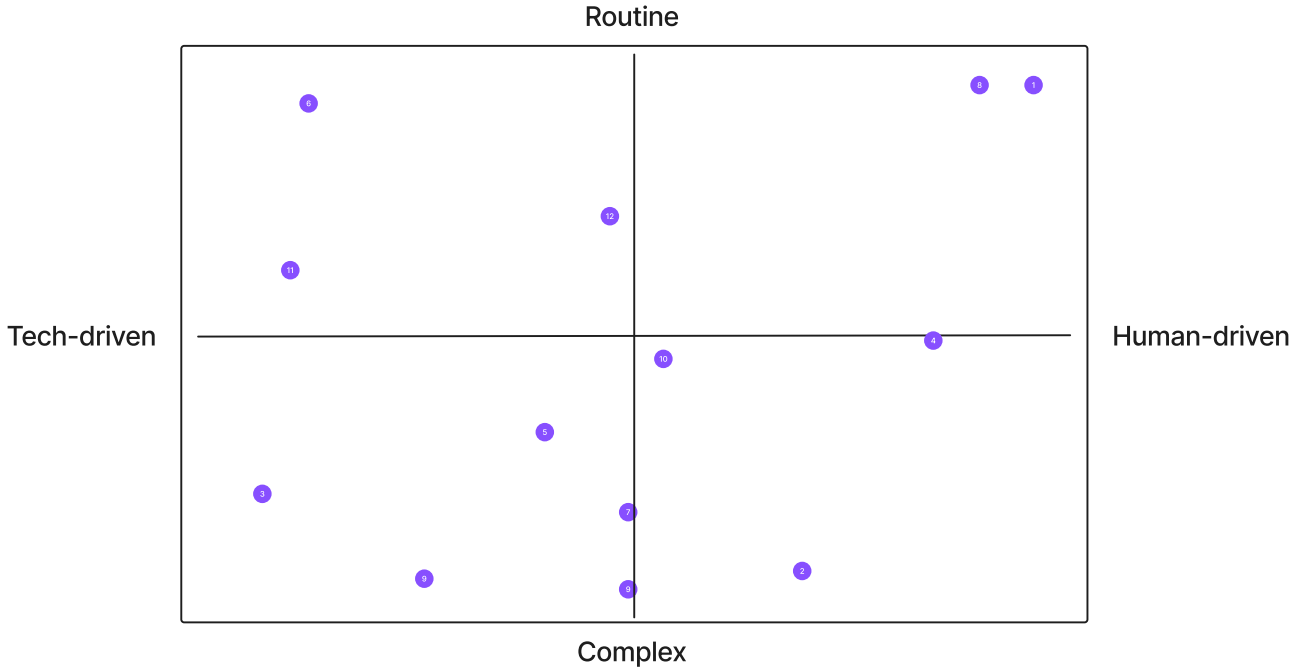


Fig. 11 Clustering of Digital Fuldmagt + Digital Post Service Operations

5.1.2 Understanding the Target Group

As established, this project was conducted in collaboration with Copenhagen Municipality's Citizen Service Development (CSD) department. Given this partnership, CSD naturally served as the primary target group. This project specifically targeted the subset of CSD employees involved in the design, development, and deployment of AI-enabled services. They are henceforth referred to as *service creators*.

5.1.2.1 Observations

To complement the insights from front-end employees, I explored how the service creators at CSD mediated the introduction of AI-enabled tools, specifically the AI-searchbot, to two key stakeholder groups: the front-end employees, who would use the tool, and upper management, who acted as sponsors and decision-makers for the tool. The aim was to understand the challenges, expectations, and concerns raised by each group in relation to the tool's adoption.

In order to do this, I observed two separate workshops, each lasting two hours and facilitated by the service creators during the early rollout phase. Around 15 participants took part in each session. I was a silent spectator during these workshops, observing the interactions between the workshop participants and facilitators, and noting important dialogue between the two. Following the workshops, I reviewed my notes and highlighted key moments of the observations.

Visual documentation of the workshops: (L) with the upper management and (R) with the front-end employees

The first workshop involved upper management and team leads from departments where the AI-searchbot was to be deployed. Its purpose was to let participants explore the tool before discussing strategies for introducing it to their teams. After a short demonstration, participants tested the searchbot and then worked in groups to generate ideas. Many recognised its potential, though some limitations became apparent. For example, the tool lacked summarised query histories - a feature some had expected based on experiences with systems like ChatGPT. This limitation reflected the project's budgetary constraints. More importantly, a few team leads struggled to see how the tool could be introduced to teams that did not currently use the existing internal search engine. This highlighted a wider challenge: building trust and encouraging adoption in teams with little prior engagement. Nonetheless, several practical suggestions were offered, such as appointing ambassadors and giving employees the option to continue using the old system alongside the new one.

Visual documentation of the various adoption strategies discussed by upper management

The second workshop involved employees from various front-facing teams. These participants were asked to test the AI-searchbot's reliability by bringing ten frequently asked questions and inputting each one ten times using identical wording. Questions had to be full W-questions, not keywords. For each correct or helpful response, participants gave a thumbs-up on the platform. If a response was incorrect, partially correct, or unclear, they gave a thumbs-down and added a written explanation. A success threshold of 80% positive feedback was set.

This session revealed several practical and behavioural challenges. Some employees resisted the need to write full questions, as they were used to searching by keywords. Participants varied in how closely they followed instructions, with some repeating each question exactly and checking every answer carefully, while others were less structured. Even among experienced employees, there was uncertainty about what counted as a correct or incorrect response. This led to confusion when giving feedback. A further concern was raised about what happens when an employee unfamiliar with a service area receives a wrong answer. In such cases, they may not realise the response is incorrect and may lack time to verify it, especially during live citizen interactions. As a result, the workshop required close support from the service creators, who guided participants throughout.

Together, these observations showed both the opportunities and limitations of introducing the AI-searchbot. While upper management expressed interest and saw potential in the tool, front end employees needed time, support, and reassurance to engage with it. The service creators played a central role in managing these differences. Acting as mediators, they balanced user needs, technical challenges, and project constraints. Although they aimed to involve end-users more in the process, tight timelines and limited resources made this difficult. Furthermore, these observations demonstrate that adoption and trust are shaped not only by the tool's technical performance, but also by how well it aligns with existing work cultures and behavioural patterns.

As the target group for this project, the service creators at CSD are uniquely positioned to shape such strategies. They act as critical mediators between end-users and organisational objectives, and their role in translating technical possibilities into usable, trusted tools will be central to the success of the AI-searchbot rollout.

5.1.2.2 Interviews

The next step in the process was to understand and explore the experiences of service creators and designers who have been involved in the development of AI-enabled services. The aim was to gain insight into the processes, challenges, and reflections that shape their work and inform the design of such technologies in both public and private sector contexts.

To achieve this, I conducted three semi-structured interviews (Adams, 2015): two with service creators from external organisations, and one with a service creator from CSD. One interview also included a software engineer, who had recently completed a master's thesis on a related topic. All participants were selected through professional networks based on their practical experience with AI-enabled services. In the interest of transparency and ethics, participants were briefed in advance, received an overview of the project and problem area, and signed informed consent forms (see Appendix C). Anonymity was guaranteed, and data handling adhered to GDPR regulations. Audio recordings and transcripts were stored locally and deleted following the project's completion (see Appendix C). The interviews were guided by a flexible interview protocol structured around core theoretical themes relevant to the project (see Appendix C). This allowed for emergent topics to be explored during the conversation, capturing a richer account of each participant's experience.

The first interview was with a service designer from a private organisation with experience in three AI-enabled projects across different European countries. The second interview was with a service designer and software developer currently working on an AI-enabled service in a large private organisation. The last interview was with a service creator from CSD involved in the AI-searchbot project, discussing their end-to-end involvement in the design and development of the tool.

Visual documentation of the second interview

Following each interview, the captured data was transcribed and thematically analysed. I organised insights using digital workspace tools, identifying common patterns and themes that cut across the interviews.

The interviews revealed several key insights. First, the introduction of AI-enabled services is fundamentally altering the role of service creators. Beyond their traditional responsibilities (such as needs assessment, service design, and stakeholder coordination), they are now required to engage with new domains of knowledge and practice. The interviews revealed how service creators must increasingly act as intermediaries between technical teams (e.g., data scientists, developers) and end users (e.g., front-end employees), translating between user needs, ethical considerations, and the affordances and limitations of the AI system. This requires not only technical literacy but also soft skills like facilitation, communication, and trust-building. Service creators are no longer simply delivering a finalised product but are engaged in a continuous process of calibration and maintenance. They must anticipate misuses, support users post-implementation, and iterate based on feedback. In this sense, their role is less about linear project delivery and more about stewardship.

Another theme that emerged across the interviews was the evolving understanding of what AI can and cannot do. Several interviewees noted that expectations at the start of a project were often inflated. AI was initially framed as a transformative solution, sometimes perceived as capable of replacing significant elements of human decision-making or labour. However, over time, service creators encountered the practical limitations of the technology. These included issues like data bias, lack of explainability, rigidity in handling edge cases, and high resource demands for training and deployment. This process of "learning the limits" often reshaped the service creators' view of the technology from a disruptive solution to a context-dependent tool that must be carefully managed. It also revealed the asymmetry between hype and operational reality. The interviews suggest that part of the design process becomes an exercise in expectation management for themselves, their teams, and their stakeholders. As such, AI-enabled service development is not just about technical implementation, but also about ongoing reflexivity, humility, and strategic adaptation to the technology's evolving capabilities.

Finally, the interviews revealed that trust in AI-enabled services cannot be reduced to a simple question of reliability or accuracy. Instead, trust emerged as a layered and dynamic construct shaped by multiple interrelated dimensions, such as technical, institutional, cultural, and interpersonal (Afroogh et al., 2024; Siau & Wang, 2018). On a technical level, trust is influenced by the system's performance, predictability, and transparency. On an institutional level, trust is underpinned by the public's confidence in the organisations deploying the technology. In countries like Denmark, where there

is a high baseline trust in public institutions, this can both enable and complicate AI adoption (Denmark Tops Europe in AI Adoption, 2025; Lorenzen, 2024). Culturally, trust is further shaped by shared values and norms regarding technology, expertise, and the role of the public sector. The interpersonal aspect is also significant: how employees speak about and engage with the system can influence how citizens perceive its legitimacy. Moreover, trust was not seen as something established once and for all. It is built over time, through communication, user involvement, and continuous improvement. Service creators reflected on how they had to actively manage expectations and provide support from inception of the service to after deployment. This insight challenges the assumption that trust results solely from good design. Rather, it must be continually negotiated and collectively maintained.

Critically reflecting on this exercise, several limitations and methodological considerations must be acknowledged. First, the small sample size and reliance on professional networks may have introduced selection bias, limiting the generalisability of findings. The perspectives captured reflect specific organisational and cultural contexts, which may not apply more broadly. Second, the use of thematic coding, while useful for identifying patterns, remains a subjective process influenced by my interpretive lens. Different researchers might have categorised insights differently or drawn alternative conclusions. Lastly, my dual role as researcher and professional peer may have shaped the dynamics of the interviews, potentially influencing what participants felt comfortable disclosing. Nevertheless, the insights gained through this process offered valuable, practice-oriented perspectives on the evolving role of service creators in the context of AI.

5.1.3 Stakeholder Map

To contextualise the service environment, I created a stakeholder map (Giordano et al., 2018) to capture the network of direct and indirect actors involved in developing and adopting the AI-searchbot. Commonly used in service and systems design, stakeholder mapping helps clarify the actors involved in a service, alongwith their role and level of influence in the service. To develop the map, I used a circular mapping format consisting of three concentric layers. The innermost circle includes stakeholders with direct influence on the service, such as the service creators, front-end employees, and team leads. The second circle represents stakeholders with indirect influence, including upper management, legal advisors, and other municipal departments whose work intersects with the tool. The outermost circle includs more peripheral stakeholders such as national regulatory bodies and external service providers.

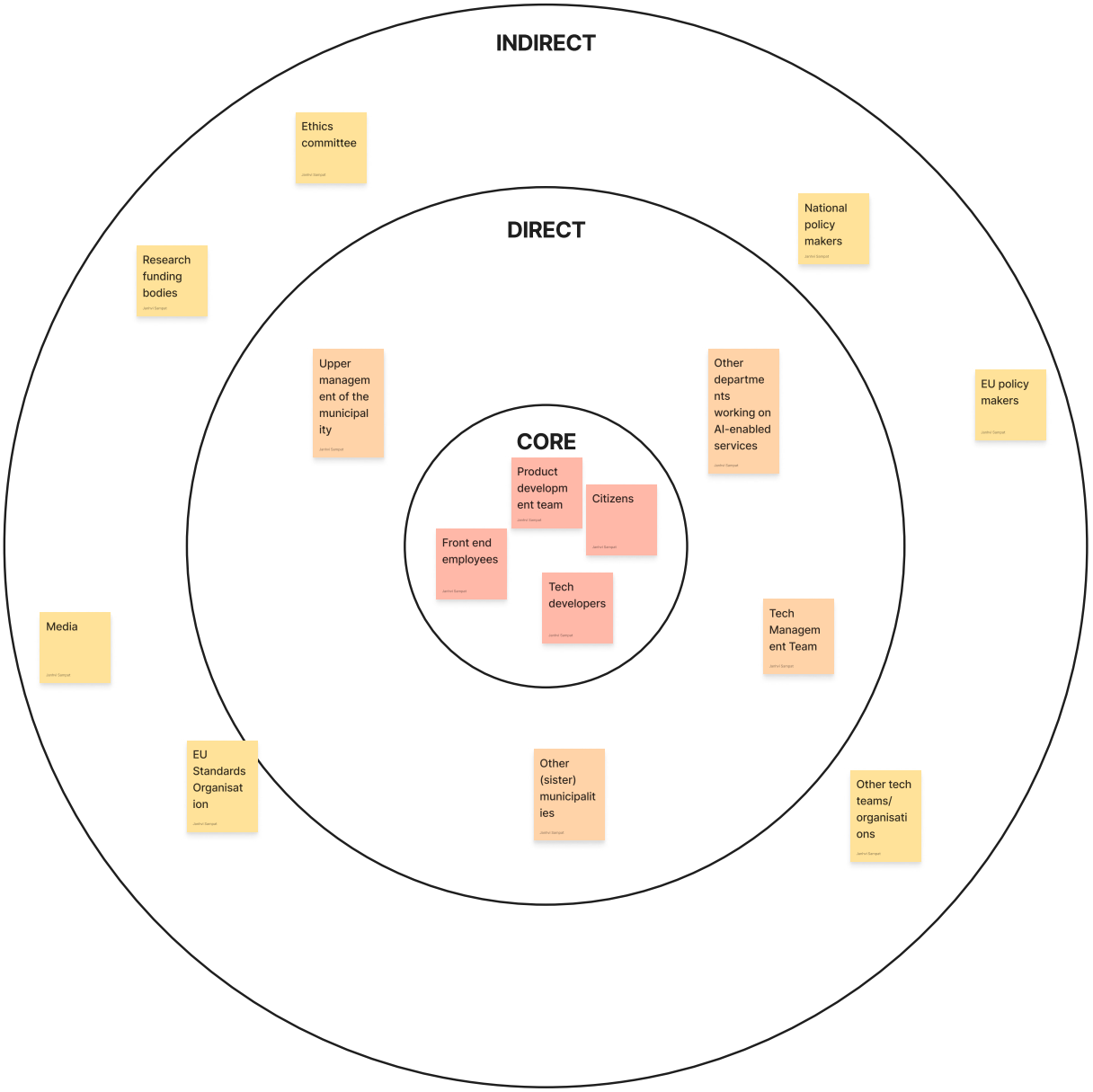
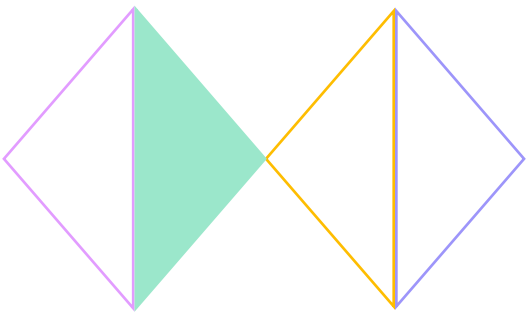


Fig. 12 Stakeholder Map

This method provided a structured overview of who was involved, in what capacity, and with what level of proximity to the service. Given the scale and complexity of the municipality, this exercise did not yield a complete picture. It was, however, instrumental in setting the scene for the project and highlighting the extent to which responsibilities and influence were distributed across actors. Critically, the mapping process began to surface broader questions of ownership, accountability, and decision-making authority. While these were not the original focus, they emerged through the act of mapping. The process also revealed the challenge of drawing clear boundaries between stakeholders in such a complex service ecosystem.

5.2 DEFINE



About This Phase

In the Define phase, I synthesised the findings from the Discover phase to better understand the dynamics shaping trust in AI-enabled services within CSD and within the larger systemic context. I developed analytical tools informed by existing theories - such as a Behavioural Model and a Trust Ladder - and applied established frameworks like the Iceberg Model, Personas, and Practice Theory to analyse underlying drivers, shifting practices, and trust-related behaviours across the system.

The synthesis process culminated in five key insights about trust in AI, a Systems Map to visualise the complexity of the problem landscape, and a refinement of the original research question. As these represent pivotal outcomes of this stage, they are presented in dedicated subchapters for clarity and depth.

5.2.1 Behavioural Model

To better understand how front-line employees at CSD make decisions and solve problems, I developed a behavioural model based on my field observations from the Discover phase and relevant behavioural theories. The goal was to synthesise the thought-process behind their behaviours and interactions with the available resources, such as the internal systems, colleagues, and their own experience. This could help locate leverage points (Meadows, 1999) for service creators to intervene and support more effective behaviours, particularly around the adoption of AI-enabled services.

This model is based on three interrelated theoretical concepts: dual process theory (Kahneman, 2011), heuristic decision-making (Gigerenzer & Gaissmaier, 2011; Kahneman & Tversky, 1979), and the Theory of Planned Behaviour (Ajzen, 1991). These domains were selected because they align with recurring behavioural patterns observed in the field: rapid intuitive responses, reliance on previous experience, and variation in motivation and confidence when engaging with different tools and systems.

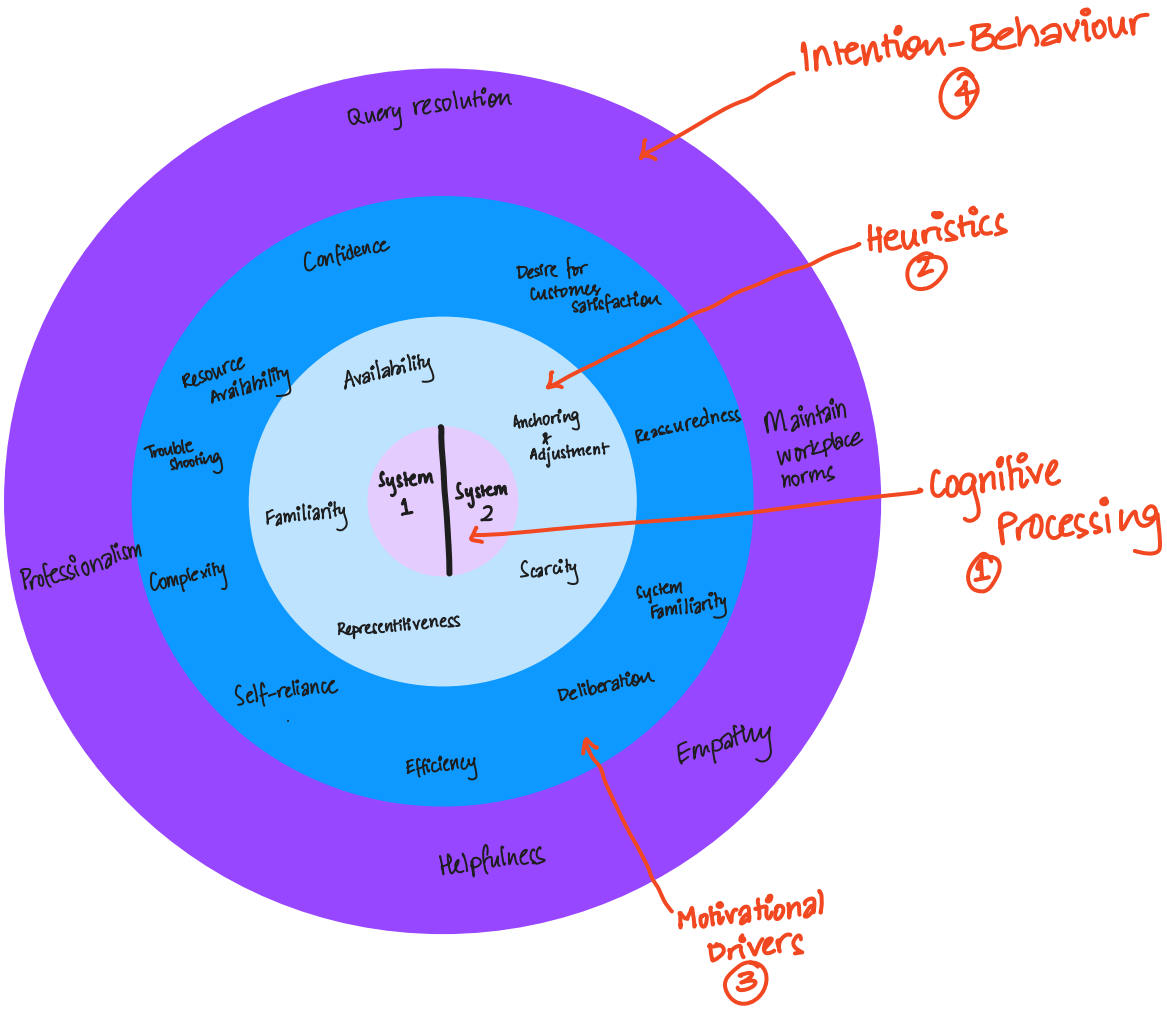


Fig. 13 Behavioural Model

The behavioural model is visualised as four concentric circles. Each layer represents a different level of behavioural influence, moving from unconscious cognitive processing to observable behaviours. While presented as discrete layers, the circles interact dynamically and continuously, with each layer shaping and being shaped by the service context.

Coginitive Processing (System 1 and System 2)

At the core of the model lies dual process theory. System 1 thinking is fast, intuitive, automatic, and typically dominant in high-paced customer service environments. System 2 thinking is slower, deliberate, and effortful, typically engaged when unfamiliar or complex queries arise (Kahneman, 2011). Both systems are activated depending on the situation.

Heuristics

The next layer comprises the heuristic strategies front-end employees employ when facing information constraints, time pressure, or ambiguity. As per observations, they frequently relied on availability heuristics (drawing on recently encountered queries), familiarity heuristics (repeating known successful behaviours), and trial-and-error methods in navigating systems and resolving customer issues (Gigerenzer & Gaissmaier, 2011). These strategies often bridged the gap between intuitive behaviours (System 1) and more reflective engagement (System 2), especially in the absence of clear guidance.

Motivational Drivers

The third layer reflects the motivational context that shapes behavioural intentions. Drawing from the Theory of Planned Behaviour (Ajzen, 1991), this layer includes: attitudes (e.g. preference for self-reliance); subjective norms (e.g. perceived expectations around service quality or collaboration); perceived behavioural control (e.g. confidence in using a system). These factors were not always consciously expressed but were evident in decision-making and task execution.

Observed Behaviours and Workplace Norms

The outermost layer encompasses the actual behaviours observed, such as knowledge sharing, empathetic communication, improvisation, and maintaining team norms. These behaviours are shaped by the inner layers, but also influence them over time. For example, repeated use of a successful heuristic may reinforce confidence and alter perceived behavioural control.

By modelling these interactions, the framework offers a nuanced understanding of how behaviours emerge among the front-end employees in Citizen Service. They are not merely isolated acts, but are situated responses shaped by cognitive, strategic, and motivational dynamics. However, I do question the direct usefulness of the model to

the core focus of this project: the service creators at CSD. On one hand, it provides a structured way to analyse the behaviours of end users, which could help CSD design more user-sensitive AI services. On the other hand, it did not directly advance the development of tools or interventions for the service creators themselves. Another limitation lies in the speculative nature of the model. It is based on my interpretation of observed behaviours, without formal validation. If the model were to be used in practice, it would benefit from refinement and validation through closer collaboration with behavioural experts or further empirical research.

5.2.2 Iceberg Model

I wanted to uncover the deeper, systemic causes underlying my observations and interactions with the various stakeholders in the Discover phase, in order to move beyond surface-level understanding of the behaviours shaping the adoption of the AI-searchbot. I believed that identifying the root causes of resistance and hesitation would allow me to design interventions that could target systemic barriers. To achieve this, I used the Iceberg Model, a common tool in systems thinking (D. H. Kim, 1999; Monat & Gannon, 2015). This model visualises how events are influenced by deeper patterns, structures, values, and unconscious beliefs. Like an iceberg, only a small portion is visible above the surface (the observable events), while most of the forces shaping these events lie hidden below.

The Iceberg Model comprises four interconnected levels that help reveal the systemic root causes behind observable behaviours:

1. **Events / Litany (surface level):** immediate, visible occurrences that happen on a day-to-day basis, such as an employee hesitating to use a new tool. They are the most apparent but offer the least insight into underlying causes.
2. **Structures & Systems (underlying patterns and trends):** When similar events recur over time, they form patterns. For example, repeated hesitation among employees across service centres may indicate a broader trend of resistance to digital tools.
3. **Worldviews & Values (normative beliefs):** These refer to the setups, workflows, or incentive systems that generate the observed patterns. They can be physical or intangible.
4. **Deep Myths (unconscious narratives):** At the deepest level lie the values, assumptions, and unconscious narratives that shape behaviour. These might include beliefs about the role of human judgement in public service or scepticism towards automation.

This exercise revealed that the resistance to AI is not solely technological. Rather, it is deeply embedded in social norms, institutional structures, cultural values, and national identity. Trust in AI is shaped by historical experience, legal safeguards, organisational workflows, and implicit beliefs about human versus machine roles. It became evident that the front end employees' preference for consulting colleagues is not irrational, but tied to the institutional legitimacy of human expertise and the lack of transparency in AI decision-making. Furthermore, CSD operates within a public sector framework that values caution, accountability, and ethical compliance - factors that often conflict with the iterative, fast-paced development typical of AI in the private sector.

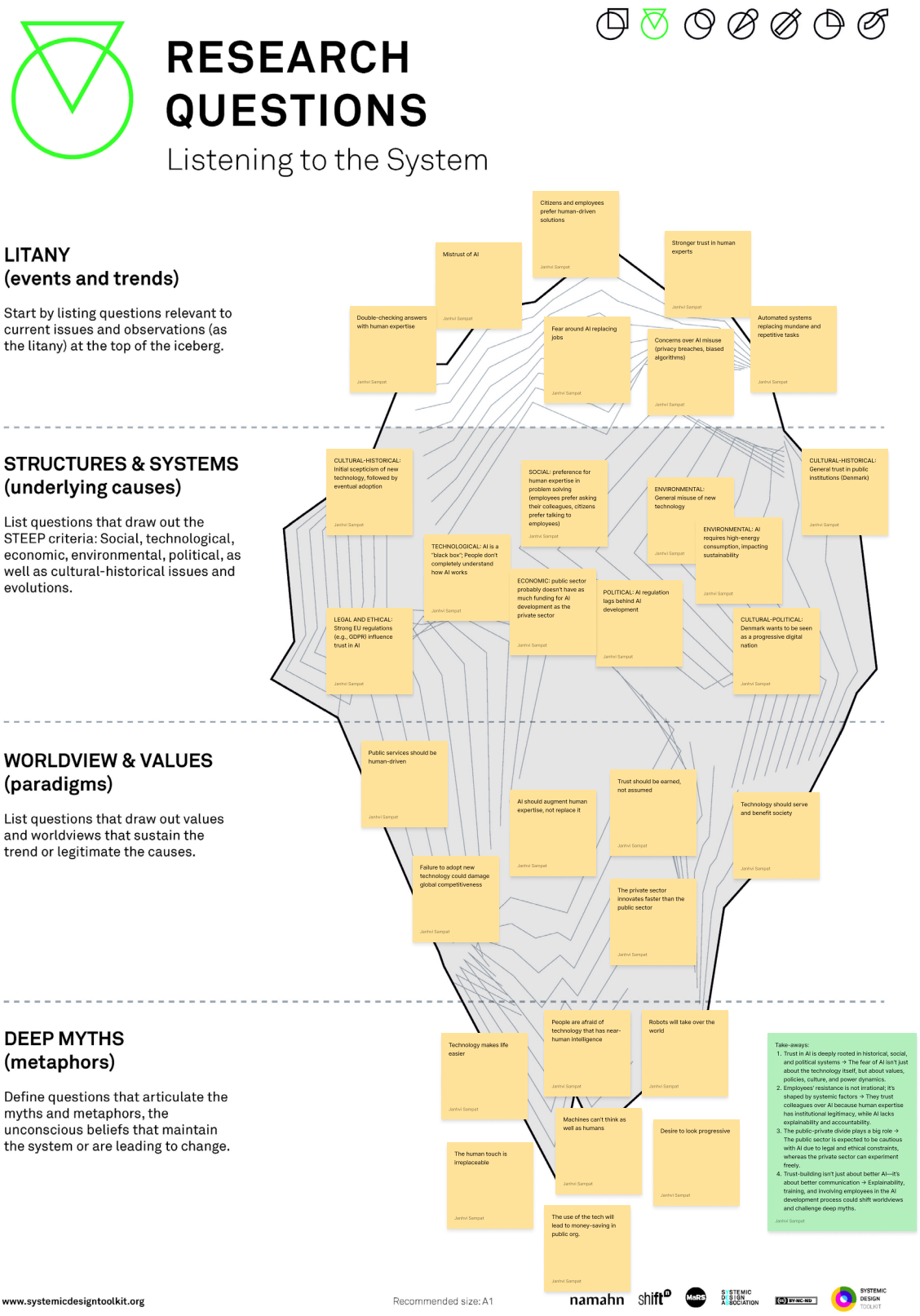


Fig. 14 Iceberg Model

Although the Iceberg Model helped uncover underlying patterns and worldviews, the exercise was not without challenges. It was difficult to separate my own assumptions from the systemic causes. There was also an overwhelming sense of complexity: acknowledging deep-rooted issues can make the path to change seem uncertain or inaccessible. Nevertheless, the model offered a valuable framework for understanding why adoption efforts might fail and where to begin designing effective, system-aware interventions.

5.2.3 Personas

5.2.3.1 Developing Personas

Personas are fictional yet evidence-informed profiles that represent archetypal users within a system or service (Miaskiewicz & Kozar, 2011). Commonly used in user-centred design, they help condense qualitative findings into relatable characters that embody distinct behaviours, needs, and attitudes. While they do not represent real individuals, they offer a structured way to keep the users' perspectives in view throughout the design process (Miaskiewicz & Kozar, 2011).

Initially, the aim was to create personas that represented different types of front end employees based on their observed interaction with digital support tools and attitudes towards AI. These were developed using the findings gathered during service operation observations and a joint workshop between service creators and front-end employees. Based on this, three personas were constructed to represent a range from AI-averse to AI-enthusiastic users. The idea was to present these personas as a resource to the CSD department, helping them remain aware of the diversity of end users when designing AI-supported services.

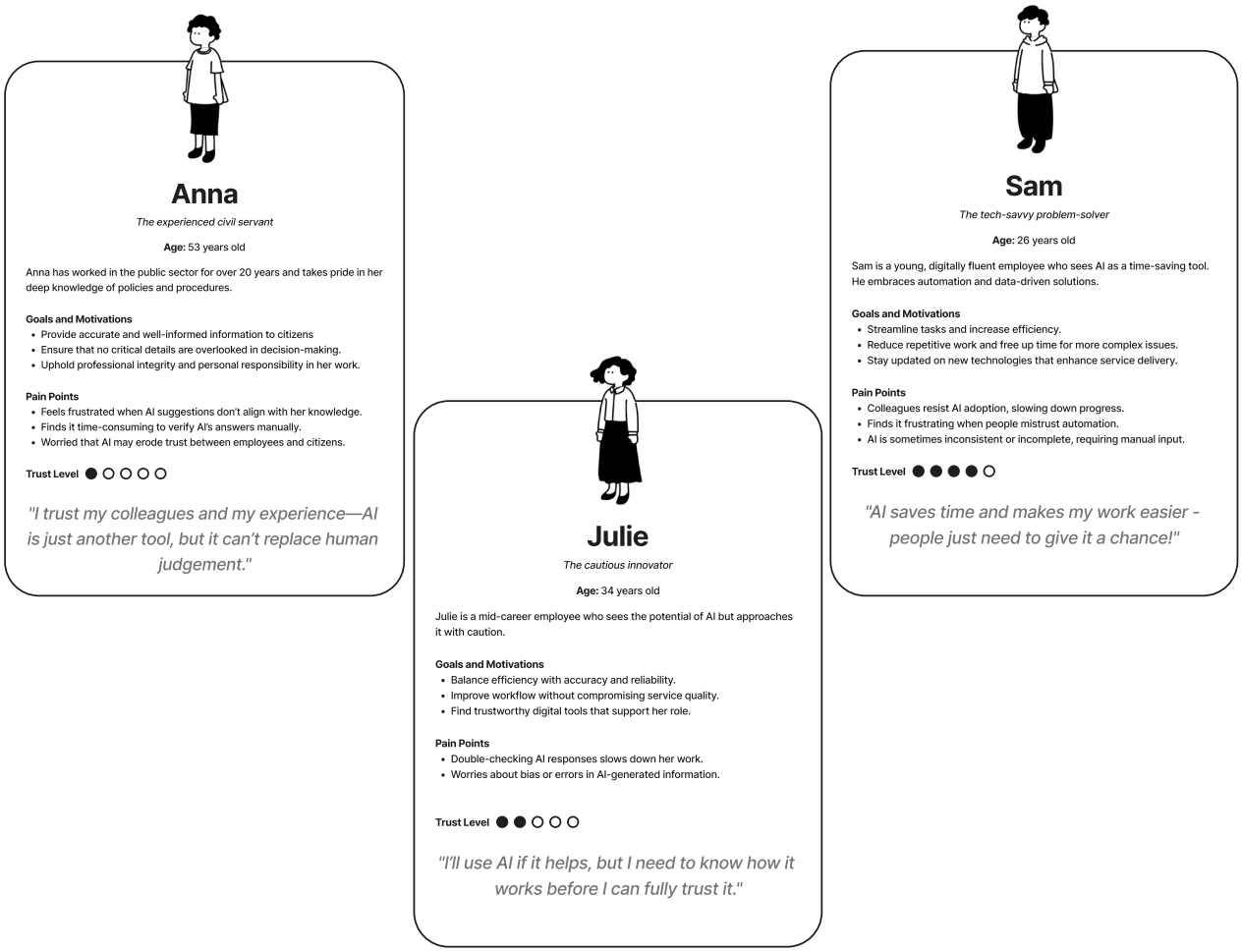


Fig. 15 Preliminary Personas

However, as the project progressed, my focus shifted more clearly to the primary users of my work: the service creators within CSD themselves. These are the individuals who conceptualise, design, and implement services and support tools for Citizen Services. Based on my observations, my impression of the CSD service creators was one of AI-positivity. This was evidenced by their proactive engagement with AI in daily tasks and initiatives such as dedicated “AI-tool days” to explore, test, and share new AI-tools with the team.

Visual documentation of an AI-Tool Day at CSD

I therefore created a simplified persona profile to represent the department’s collective approach to AI. Moreover, I created a similar simplified persona that represented the upper management’s attitude towards AI. Given the organisational push towards more AI-enabled services and the enthusiasm for AI-searchbot observed during workshop in the Discover phase, I assumed them to be more accepting and open to AI, in general. However, I do lack a more nuanced understanding of their attitudes towards AI and their general familiarity and use of the technology.

AI Advocate

Embraces AI, sees it as transformative, actively promotes its adoption. Many CSD employees fall into this category.

Trust Level: ●●●●◐

AI Pragmatist

Open to AI if it demonstrably improves workflows, cautious but supportive. Upper management and some front-end employees possibly fall into this category.

Trust Level: ●●●◐○

AI Doubter

Sceptical of AI's usefulness or fairness; needs strong proof to gain trust. Some front-end employees might fall into this category.

Trust Level: ●●○○○

AI Resistor

Distrusts AI and perceives it as disruptive or threatening to their role. Some front-end employees might fall into this category.

Trust Level: ◐○○○○

Fig. 16 Simplified Personas (illustrations sourced from Freepik.com)

5.2.3.2 Matrix Mapping: Approach vs. Adoption

To complement the simplified personas and further contextualise stakeholder attitudes, I developed a scenario matrix to map the positioning of the three primary stakeholder groups, based on their current engagement with and orientation towards AI tools (Rhydderch, 2017). The matrix served as a visual and conceptual tool to synthesise observations and highlight relational dynamics across the stakeholder landscape.

The horizontal axis of the matrix represented attitude towards AI, ranging from sceptical to enthusiastic, while the vertical axis indicated actual engagement with AI tools in practice, from low to high. By plotting the stakeholder groups along these axes, it became easier to identify patterns and potential tensions or alignments.

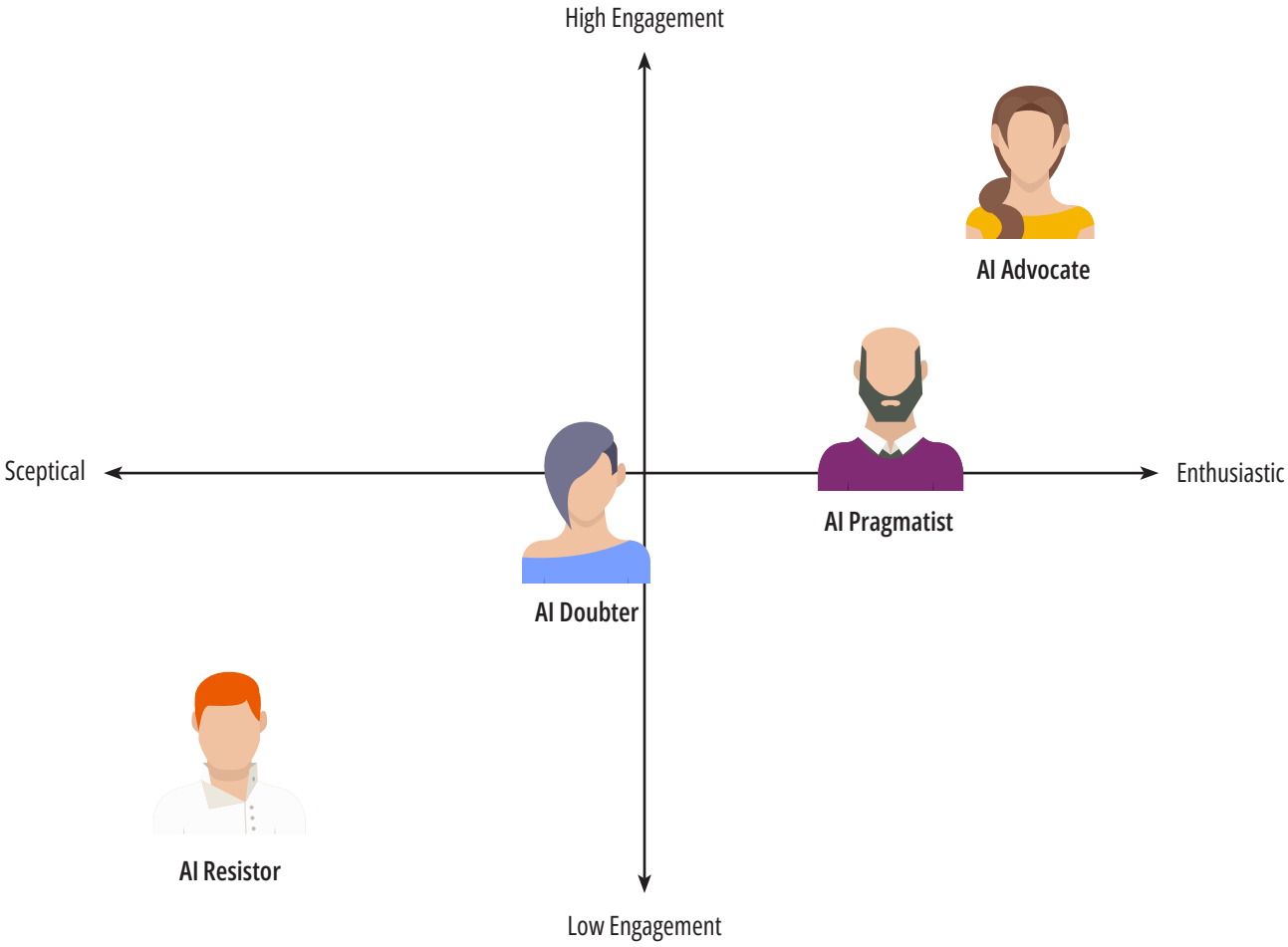


Fig. 17 Persona Mapping

The outcome of this exercise indicated that the majority of stakeholder groups, namely the CSD service creators and upper management, are currently AI-positive, both in terms of mindset and, at least for the CSD team, practical engagement. In contrast, front-end employees exhibited a wider spread, with some individuals showing high levels of scepticism and low engagement, while others were more open or enthusiastic. This visualisation, while simplified, helped communicate the diversity of readiness and

identify where support or adaptation might be most needed when designing AI-supported interventions.

Despite critiques that personas can oversimplify complexity or rely on assumptions if not rigorously grounded (Matthews et al., 2012; Salminen et al., 2018), I chose to use them for two reasons: (i) they primarily served as an internal sensemaking tool to understand and convey the various behavioural patterns and attitudes towards AI observed during fieldwork, and (ii) they were intended to function as a communication tool for the CSD team. This role, however, remained more aspirational than realised during the project.

5.2.4 Trust Ladder

As part of synthesising how trust in AI technologies is built, I developed a conceptual model, hereby referred to as the Trust Ladder. The purpose of building this model was to explore how trust in AI might grow over time. I wanted to see if there were different steps or moments where design could help support or influence that trust-building process.

The model was loosely inspired by Pine and Gilmore's (1999) Progression of Economic Value, which describes how organisations move from delivering commodities to creating transformative experiences. Additionally, it was inspired by theories of incremental innovation (Christensen, 1997; Markides, 2006), which emphasise gradual, cumulative change. Drawing from these frameworks, I attempted to model trust in AI as a staged progression - beginning with user supervision and moving toward full autonomy - mirroring how responsibility is incrementally delegated in trusted human relationships, for example a relationship between parent and child.

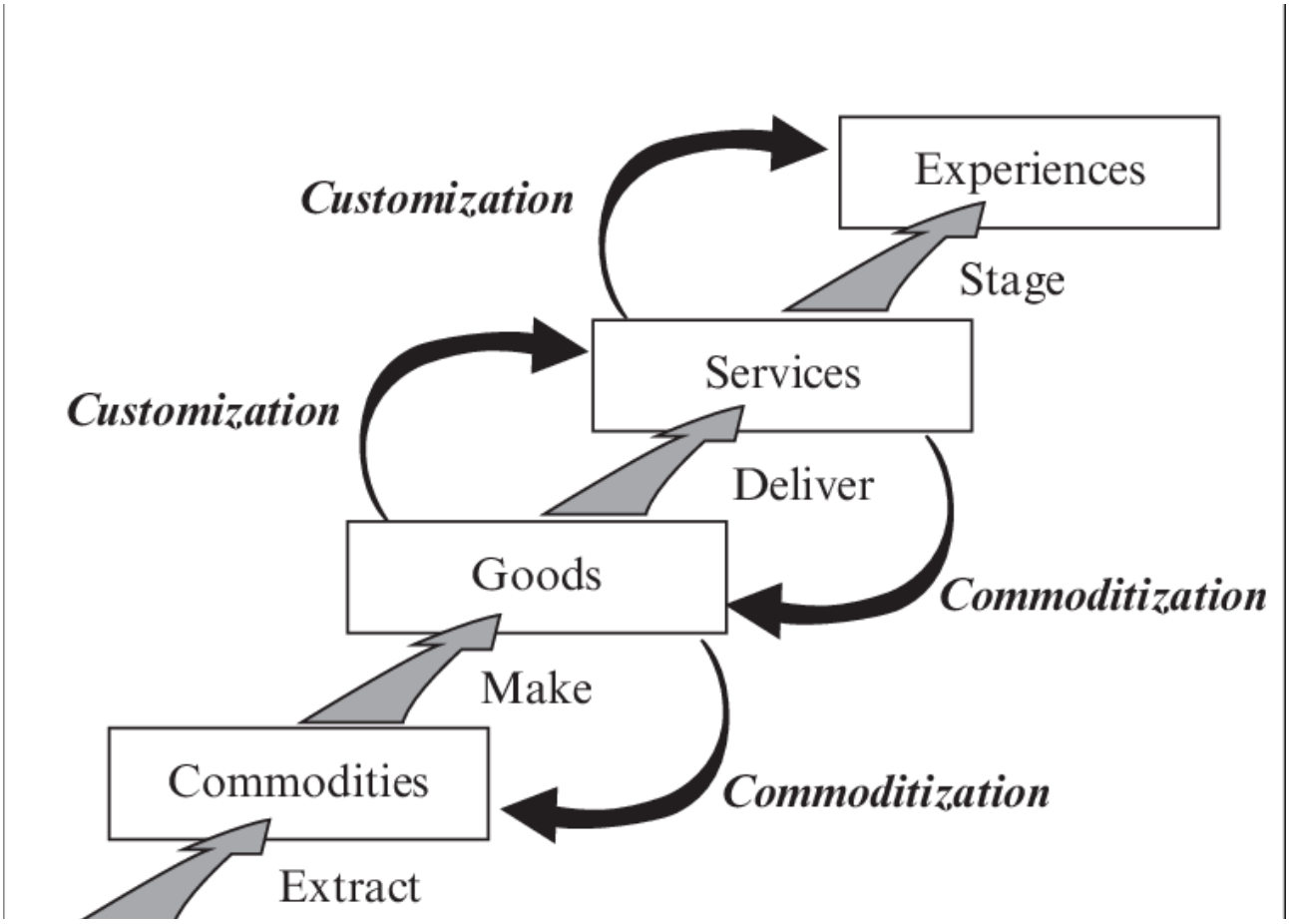


Fig. 18 Pine and Gilmore's modified Progression of Economic Value Model
Pine II, B. & Gilmore, James. (2013). The experience economy: past, present and future. 10.4337/9781781004227.00007.

This resulted in a five-stage conceptual model. Each stage was developed by reflecting on behavioural cues and dependencies that commonly appear in trust relationships, abstracted here for application to human–AI interaction.

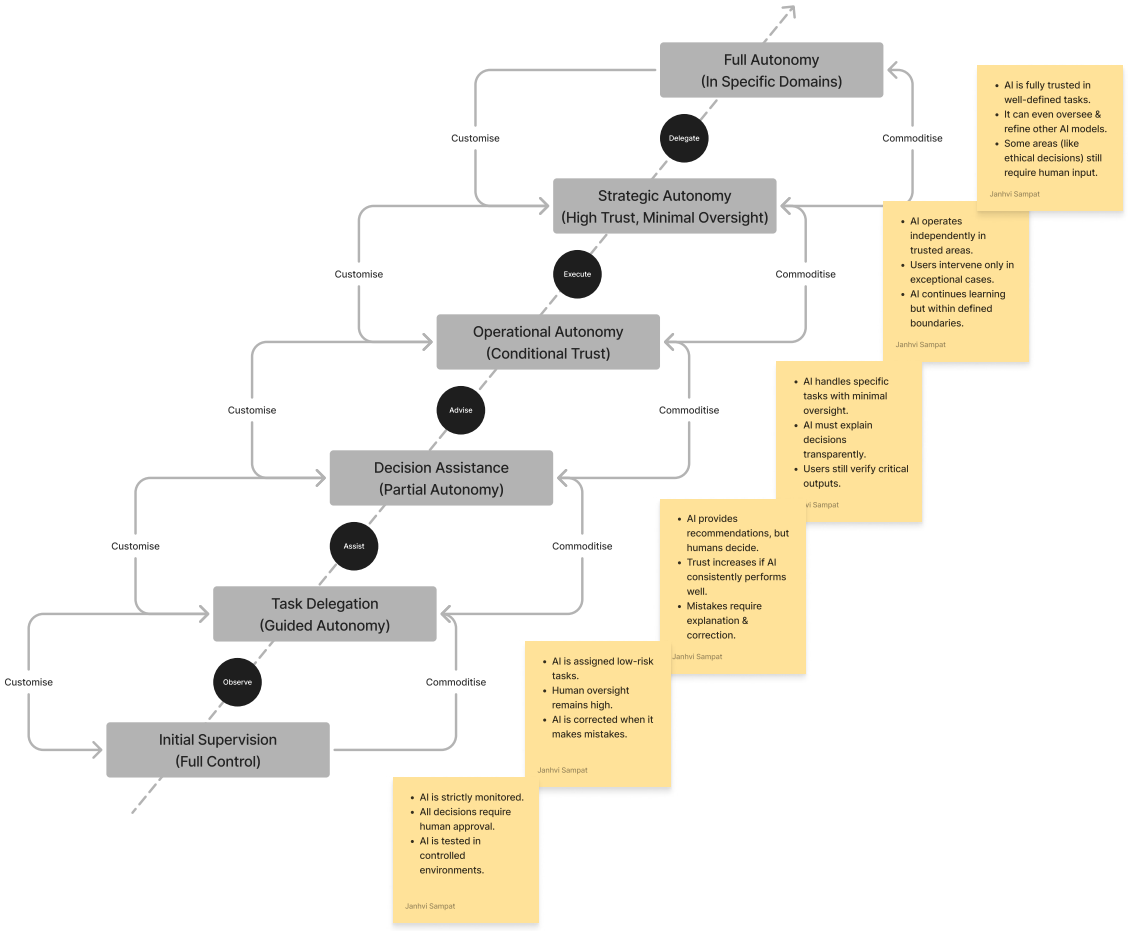


Fig. 19 Trust Ladder

This resulted in a five-stage conceptual model. Each stage was developed by reflecting on behavioural cues and dependencies that commonly appear in trust relationships, abstracted here for application to human–AI interaction.

5.2.4.1 Initial Supervision → Observe → Task Delegation (Guided Autonomy)

At this early stage, the user monitors the system closely, retaining full control. The AI may provide suggestions or basic functions, but these are vetted or ignored at the user's discretion. The goal here is to observe reliability over time.
For example: a user testing a grammar-checking tool, such as Grammarly, for reliability and accuracy (Grammarly, n.d.).

5.2.4.2 Task Delegation → Assist → Decision Assistance (Partial Autonomy)

Following positive observation, the user begins to allow the AI to assist in limited, low-risk tasks. The user retains authority but accepts minor interventions. The AI's outputs begin to influence action, though ultimate responsibility still lies with the user.
For example: A driver using Tesla's Autopilot on highways but keeps hands on the wheel and constantly monitors performance (Autopilot and Full Self-Driving (Supervised) | Tesla Support, n.d.).

5.2.4.3 Decision Assistance → Advise → Operational Autonomy (Conditional Trust)

Trust begins to solidify. The AI is allowed to operate more independently within predefined constraints. Its advice is no longer supplementary but plays a central role in decision-making. However, oversight remains, and trust is conditional on consistent performance.
For example: A financial analyst uses AI-based forecasting tools like Bloomberg Terminal's ML modules ('Bloomberg Terminal', n.d.). The system makes detailed projections and even suggests actions, but decisions are still reviewed and approved manually.

5.2.4.3 Operational Autonomy → Execute → Strategic Autonomy (High Trust, Minimal Oversight)

At this stage, users begin to rely on the AI for complex or strategic decisions, stepping back from routine oversight. Trust has become routinised and embedded. The AI is assumed to understand the context and act in alignment with user expectations.
For example: In a smart warehouse, an AI-powered inventory system manages restocking and logistics with minimal human intervention (Kesari, n.d.).

5.2.4.3 Strategic Autonomy → Delegate → Full Autonomy (in Specific Domains)

Full trust is granted within a bounded domain. The user delegates decision-making and action without regular intervention. The AI is treated as a competent actor, expected to manage complexity and uncertainty within its scope.
For example: In aircraft autopilot systems during long-haul flights, pilots hand over full control to the system for hours at a time (What Is an Aircraft Autopilot?, 2024).

While the ladder served as a reflective tool, its usefulness lay not in its prescriptive accuracy, but in what it revealed during the modelling process. Several limitations of the process became evident through critical reflection. Firstly, I realised that trust in AI does not evolve in the same emotional or relational way as it does in close human relationships, such as between a parent and child. A more fitting analogy might be a professional context; for instance, onboarding a new colleague. While still a human comparison, this framing focuses less on emotional bonding and more on how trust is

formed through demonstrated competence, reliability, and shared context over time. In my opinion, this perspective aligns better with how users evaluate and gradually rely on AI systems in practice. Second, the ladder's linear structure oversimplified the trust-building process, which is inherently dynamic and influenced by numerous contextual, cultural, and interpersonal factors. Afroogh et al. (2024) argue that trust in AI is not a singular or fixed state but a composite of multiple technical and non-technical elements, such as: safety, fairness, interpretability, and social alignment, which evolve over time and interaction. Viewing trust as an iterative loop influenced by multiple factors, rather than a linear climb, may provide a more accurate representation. Finally, while the model did not produce a framework that could be directly applied or validated in practice, it supported an important shift in understanding: trust-building is often shaped by small, often invisible acts that accumulate over time. This insight informed subsequent design choices, where reinforcing trust became a matter of enabling subtle but consistent user–AI interactions.

5.2.5 Analysing Shifting Practices

To understand how the introduction of the AI-searchbot was changing daily work routines, I analysed the practices of two key stakeholder groups: service creators and front-end employees. The goal was to explore how their workflows were shifting as they moved from traditional service delivery to AI-enabled services. I used a simplified version of practice theory for this analysis (Kuijer, 2014; McColl-Kennedy et al., 2015).

Practice theory is a social theory framework that focuses on what people actually do in their daily work. It sees behaviour not just as a result of individual decisions or external structures, but as part of routines shaped by tools, skills, and cultural norms. A commonly used model within this theory breaks practices into three key elements: materials (the tools and technologies used), competencies (the skills and knowledge needed), and meanings (the values and assumptions that guide behaviour) (McColl-Kennedy et al., 2015).

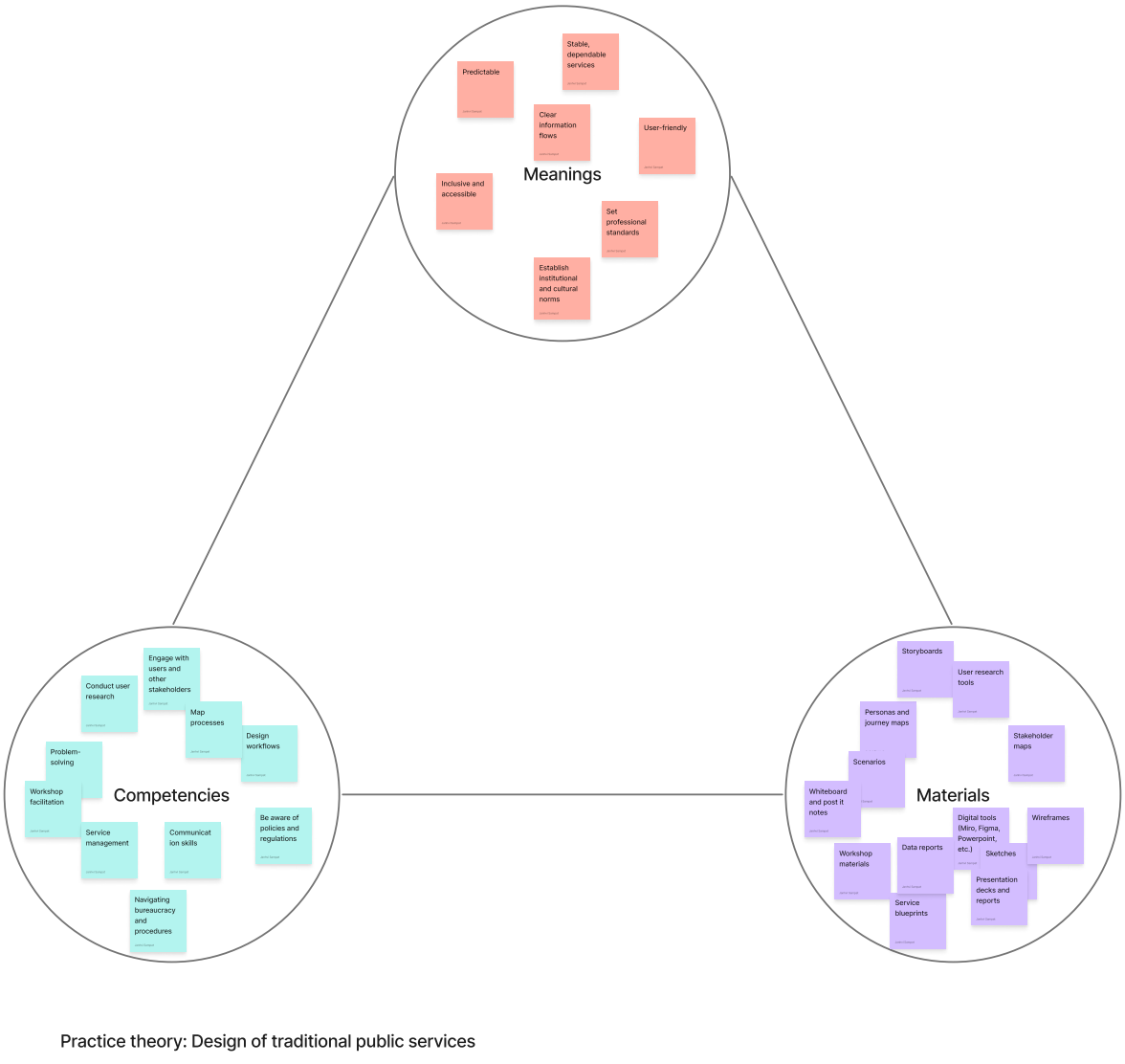
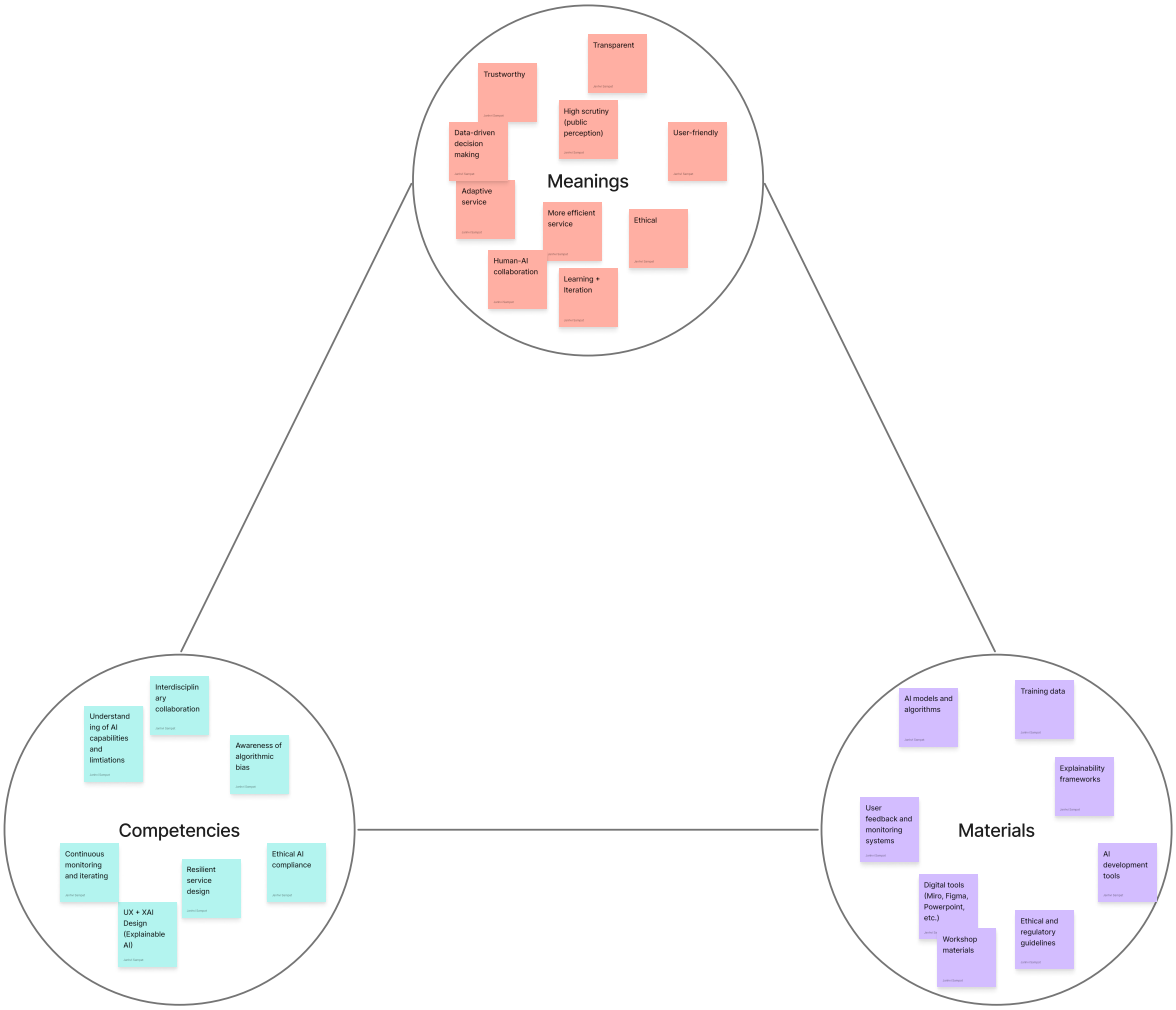


Fig. 20 Practice Theory Model for service creators designing digital or traditional services



Practice theory: Design of AI-enabled public services

Fig. 21 Practice Theory Model for service creators designing AI-enabled services

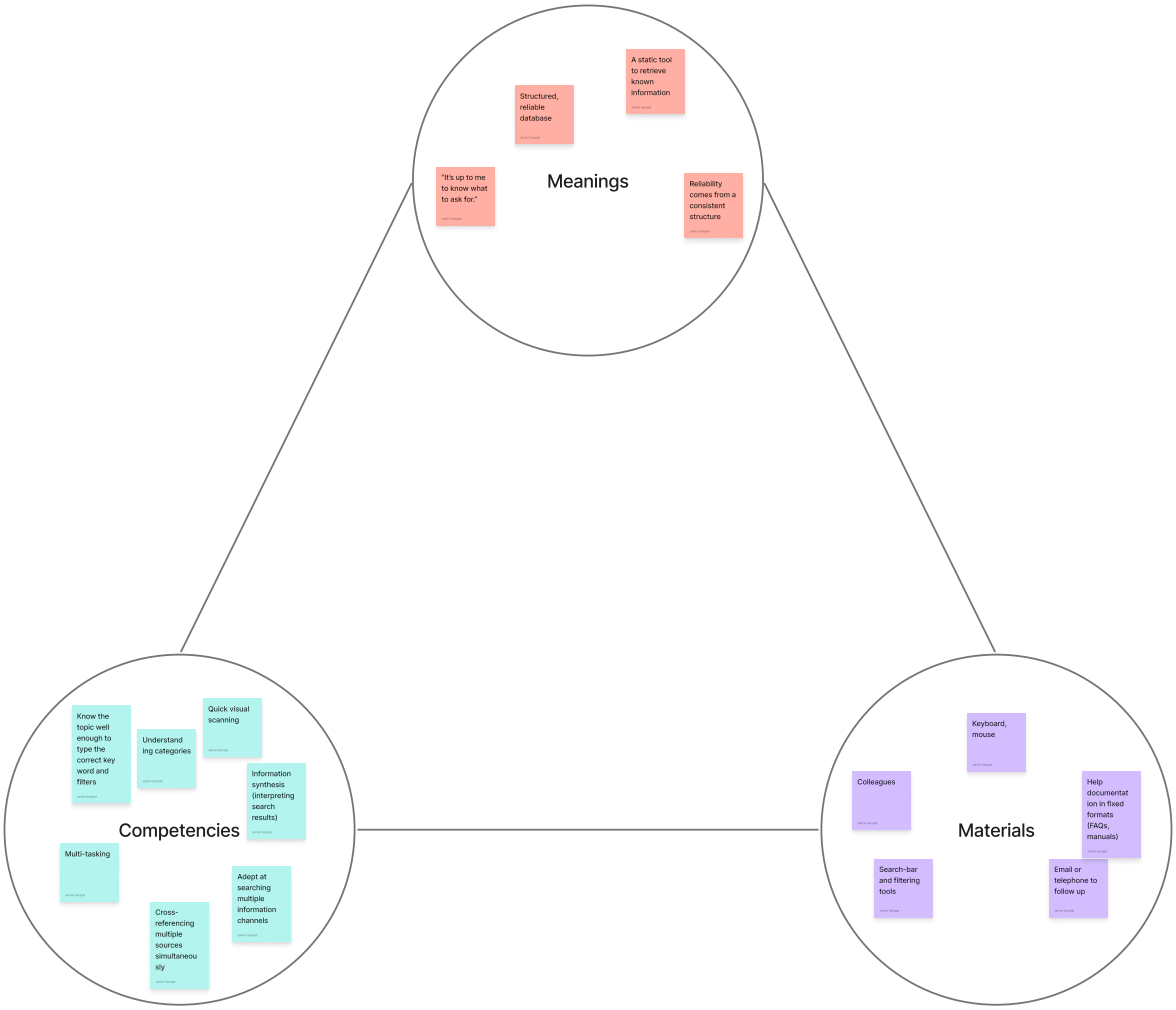


Fig. 22 Practice Theory Model for front-end employees using digital or traditional services

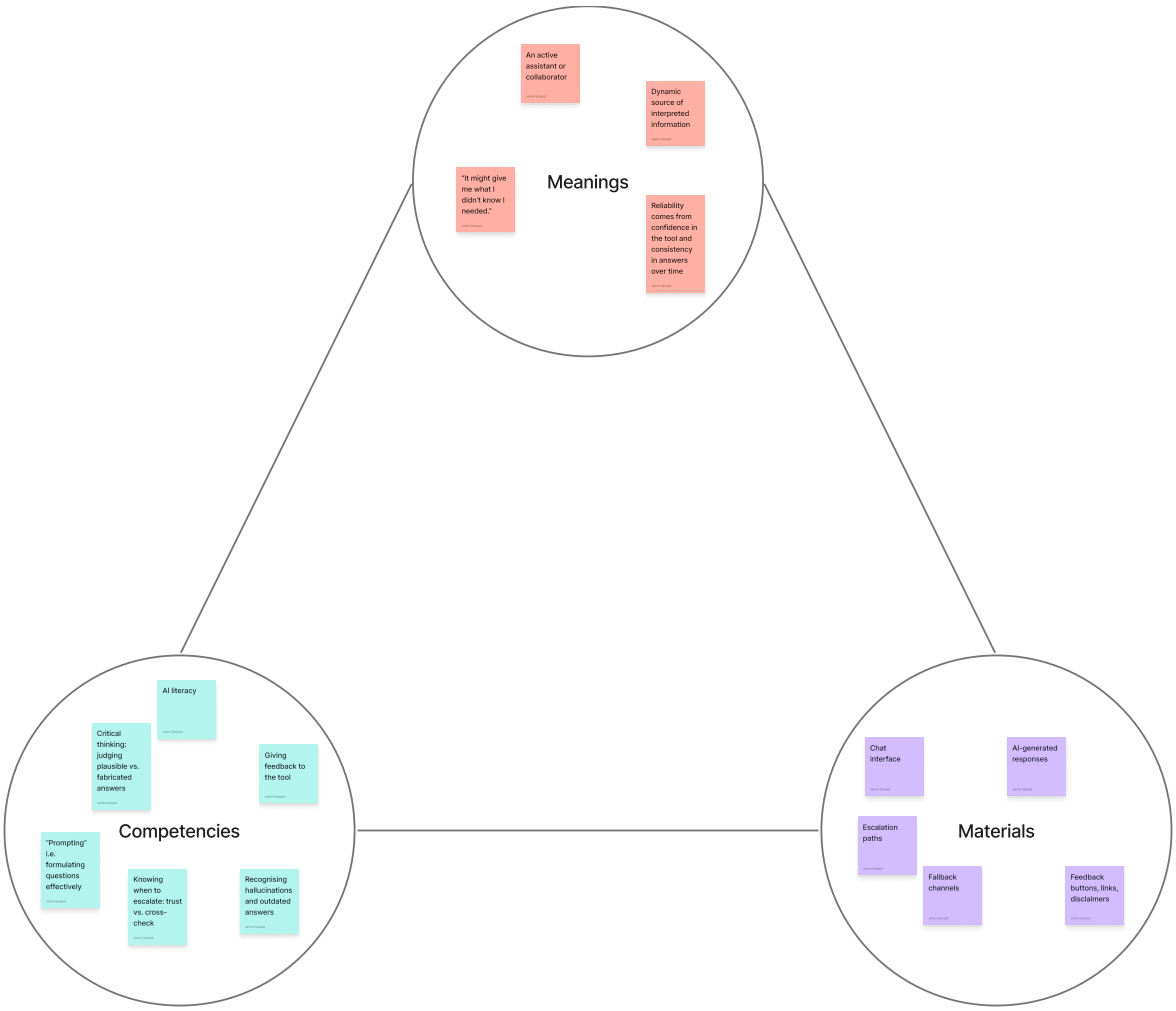


Fig. 23 Practice Theory Model for front-end employees using AI-enabled services

The analysis showed that the shift to AI-enabled services brought noticeable changes for both groups. For service creators, their work became more interdisciplinary and required new knowledge about AI's strengths and limitations. They also took on a greater role in supporting front-end employees through training and guidance. Their job now involved not just designing services, but also maintaining a feedback loop to help improve the AI tool over time.

For front-end employees, the changes were also significant. They had to get used to new ways of searching for information (moving from short keywords to full prompts) and learn how to give structured feedback on the tool's responses. They also had to rethink how they interacted with the system, and see the AI not just as a database, but as a kind of digital colleague that supports their work. The exercise also showed that the introduction of AI influenced how work was understood and valued. Service creators now had to shift their role from simply delivering a solution to helping users engage with and trust a new kind of tool.

This approach was useful to this study because it broadened my perspective on the introduction of AI from not only a technical change, but something that affects how people work and think. At the same time, there were limitations to this exercise. Breaking real-life work into the three elements of practice theory involved interpretation, and some overlap between categories was unavoidable. What I saw as a material, someone else might have considered a competency or value. This subjectivity is a known issue in practice theory (Shove et al., 2012). In addition, practice theory usually assumes that change happens gradually (Reckwitz, 2002), but the introduction of AI can lead to much faster shifts, making it harder to apply in a straightforward way. Still, the model helped me better understand how introducing an AI tool was not just about using new technology, but about changing how people work, what they need to know, and what they see as valuable or trustworthy in their roles.

5.3 INSIGHTS

5.3.1 The Mismatch Between Trust Development and the Tech Development

The first key insight that emerged during the project was the misalignment between the pace of technological innovation and the development of trust (Joseph, 2010; Public Attitudes toward New Technologies, n.d.; Why Do People Resist New Technologies?, 2016; Stilgoe, 2023). Particularly in the context of AI-enabled public sector services, technological progress often far outpaces users' ability to understand, adapt to, and develop trust in these new tools. Each new generation of technology introduces not only new capabilities, but also new risks, expectations, and behavioural shifts. However, the mechanisms by which trust is cultivated remain slow, implicit, or reactive (Mollering, 2006; Schilke & Cook, 2013). This creates a recurring gap: while AI tools advance rapidly through technical milestones, the trust required for their adoption lags behind. Trust, unlike technical functionality, cannot be shipped with a release or measured through system performance alone. It builds incrementally, shaped by transparency, reliability, and user experience over time.

This tension can also be contextualised through the technology adoption lifecycle (G. A. Moore, 1991; Rogers, 1962). While early adopters may tolerate uncertainty, the majority of users require reassurance, guidance, and demonstrated value before engaging with new systems. Trust, in this sense, becomes a prerequisite for bridging the so-called "chasm" between innovation and mainstream adoption.

This misalignment has several implications for public service design:
• **For users**, it places an unrealistic expectation to adapt to tools whose inner workings they do not understand or fully trust. Trust is eroded further by the opaque nature of many AI systems (Barredo Arrieta et al., 2020), where outputs often lack clear justification.

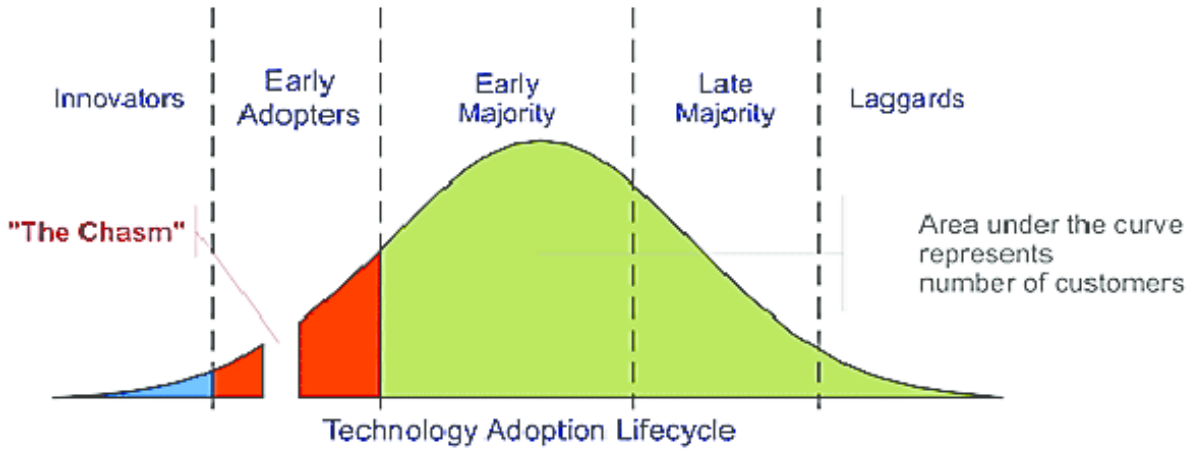


Fig. 24 Technology Adoption Lifecycle
Source: Moore, G. A., & McKenna, R. (1999). Crossing the chasm.

• **For service creators**, it means designing for trust cannot be an afterthought. They must intentionally integrate trust-building strategies into design and delivery processes.
• **For organisations**, there is a need to track trust maturity alongside technical readiness. By focusing solely on features, they may risk overlooking user hesitation or rejection.

Despite its importance, trust-building is rarely formalised in AI-related service development. Project lifecycles tend to emphasise technical deployment and delivery milestones, with few mechanisms to assess or support trust readiness. As Dietvorst et al. (2014) note, *"Although people may be willing to trust an algorithm in the absence of experience with it, seeing it perform - and almost inevitably err - will cause them to abandon it in favor of a human judge. This may occur even when people see the algorithm outperform the human."*

This insight suggests the need to shift trust from being an assumed byproduct to a core design concern. One potential approach could be to introduce formal trust checkpoints: moments during a project lifecycle where user trust, understanding, and acceptance are explicitly assessed and supported.

5.3.2 Trust Does Not Exist in a Vacuum

The second central insight that emerged in the design process was the interconnected nature of trust (Evans & Wensley, 2009; Higgins & Kruglanski, 2007) i.e. the development of trust is not an isolated event. Instead it can be considered as a dynamic, relational system of people, roles, technology, and tools (Latour, 2005). The introduction of a new entity (in this case, the AI-searchbot) alters the configuration of this system. In such a scenario, trust is often quietly redistributed and recalibrated across the network. Even a subtle disruption can have significant implications for how the system is experienced by the actors. A breakdown in one node or relationship may cascade into others, weakening the overall system. In this sense, trust is only as strong as its weakest link.

This insight is supported by sociotechnical perspectives on trust, which emphasise its situated and relational nature (Latour, 2005; Nissenbaum, 2001). Technologies introduced into organisational settings do not simply perform a function. They reconfigure relationships and redistribute agency. As trust is transferred from human actors to a machine or shared among both, it can challenge existing team norms and role expec-

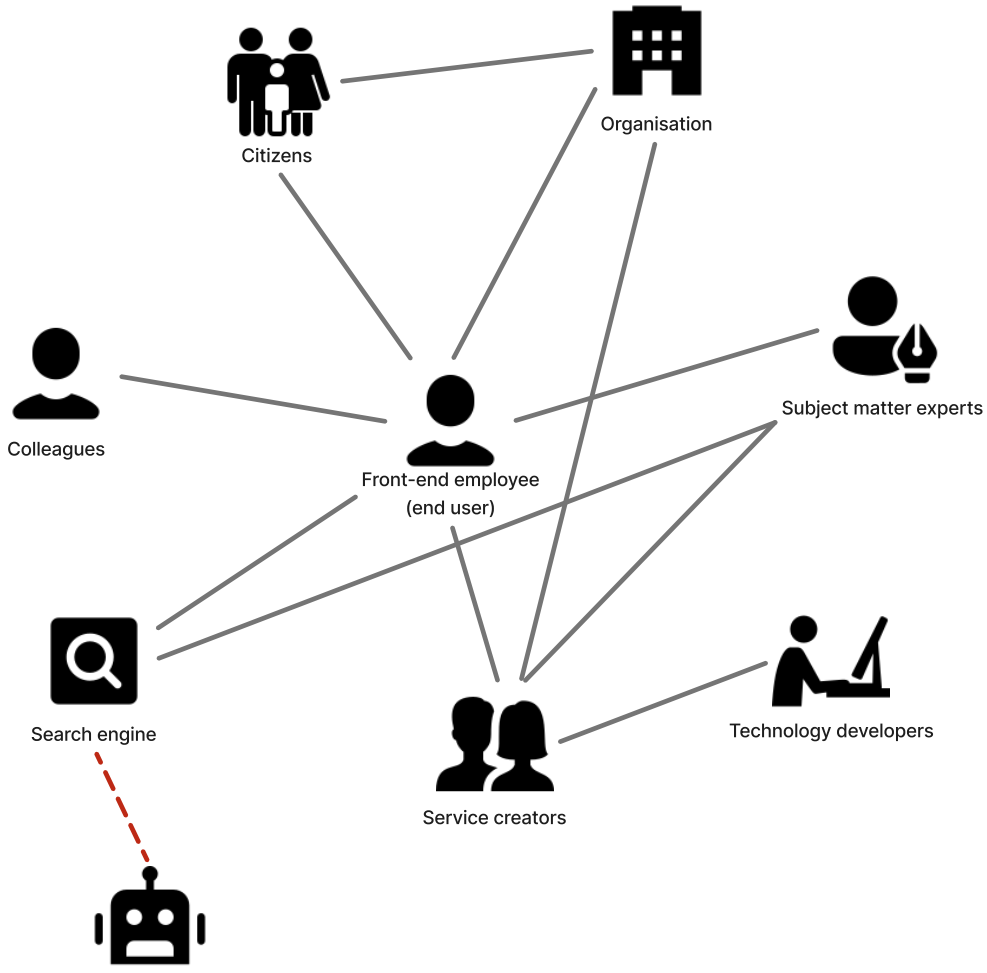


Fig. 25 Trust Network illustrating the introduction of a new entity that could cause disruption within the network

tations (Binns et al., 2018).

This shift could have several implications. Firstly, existing trust relationships can be unintentionally undermined. If the AI system fails or behaves unpredictably, it may not only lose trust itself but also reduce confidence in the system as a whole. Secondly, the introduction of AI may necessitate a redefinition of roles and responsibilities. For instance, employees may begin to rely on the system for judgement or information retrieval, altering their own perceived authority or accountability. Thirdly, the integration of AI is not only a technical process, but a social one. Therefore, social and relational dimensions of trust must be given equal weight to technical considerations in design.
Designing AI-enabled services within such a complex trust network is inherently challenging. Changes within the trust network can create ambiguity about responsibilities, and subtle tensions in team dynamics may emerge. Moreover, changes in trust patterns are often difficult to observe directly, making it hard for service creators to anticipate or respond to risks.

These findings point toward the need for trust-aware design practices that support the existing trust network and ease its disruption.

5.3.3 AI is an Organism, not a Static Tool

Unlike traditional digital systems which are governed by fixed rules, AI systems are dynamic and evolve through exposure to real-world data and use (Bennett, 1987; Driankov et al., 1996; Siau & Wang, 2018). The behaviour of these systems is emergent and unexpected, resulting in outputs and behaviours that may not have been predicted at the time of design. This introduces a sort of "trust paradox", where while users tend to expect consistency and reliability and AI inherently embodies uncertainty and change. This raises the question of whether it is necessary to redefine trust itself in these scenarios to encompass resilience and adaptability.

"I think of AI as an acronym not for artificial intelligence, but for alien intelligence. I mean alien not in the sense that it's coming from outer space, but alien in the sense that it thinks, makes decisions and processes information in a fundamentally different way than humans. It's not even organic. The most important thing to realize about AI is that it is not a tool. It's an agent. Every previous technology in history was a tool in our hands. You invent a printing press, you decide what to print. You invent an atom bomb, you decide which cities to bomb. But you invent an AI, and the AI starts to make the decisions. It starts to decide which books to print and which cities to bomb, and eventually even which new AIs to develop. So don't think about it like the previous technologies we've had in history. This is completely new."

- Yuval Noah Harari
Author

This insight uncovers several implications. One that trust must be maintained on an ongoing basis, rather than a one-time design goal. Additionally, there are certain essential mechanisms available to reinforce and maintain trust in AI-systems, such as real-time oversight and feedback options. However, a few challenges that come to light, as well. Users may struggle to understand why the AI behaves differently over time, especially without transparent feedback loops. Additionally, organisational structures may lack processes for revisiting or revising decisions as AI models evolve. These insights and understandings highlight the need to design AI-systems that evolve transparently, without disrupting established trust.

5.3.4 New Practices for Designing and Using AI-Enabled Services

The next key insight from the process was how the introduction of AI- enabled services not only adds new capabilities to an organisation, it fundamentally alters the practice of designing and using that service (Dove et al., 2017; Gaver, 2012; Long & Magerko, 2020). Traditional service design typically assumes stable rules and predictable outcomes. In contrast, designing for and using AI-enabled services call for the adoptions of new routines, behaviours, and understandings of the system. As a result, both the practice of designing and using the service are altered.

"I think the biggest issue here is that AI is such a new field, and having experience with it - well, it kind of requires that you've worked in data science or in a development department, that you're a data scientist and have worked with machine learning models..."

- CSD Service Creator

Source: Interview

The insight uncovers certain systemic shifts. It implies the need for new or adapted models, methods, and tools that account for the socio-technical complexity of AI. Moreover, certain actors and stakeholders (such as data scientists, legal experts, ethics advisors) may take on more central roles in the new ecosystems. For end-users, engaging with AI-enabled services can mean adapting to new workflows, levels of autonomy, and expectations, all of which point to the need for more support, training, and clear communication. This insight also considers a set of challenges. Designers may feel underprepared or excluded, as traditional human-centred design approaches may seem at odds with the probabilistic nature of AI systems. The values and principles of AI systems (such as optimisation, prediction, or automation) may conflict with core design values like empathy, agency, and inclusivity (Auernhammer, 2020; Chen et al., 2022). Navigating these tensions may call for not only technical upskilling, but also reflective capacity and organisational support.

Supporting both service creators and users in adapting to these shifts is essential for sustaining trust. The insight therefore raises the question of designing for - and with - the changing practices of designing and using a service.

5.3.5 The Invisible Work of Trust-Building

The building and maintenance of trust, especially in public services, entails a lot of invisible work (Star & Strauss, 1999). This trust is built incrementally through small, interpersonal efforts (such as emotional labour, ongoing user support, and informal coordination) that bridge the inevitable gaps between system design and real-world complexity. This form of invisible work is rarely accounted for, even though it is critical to building and maintaining trust. In the context of AI-enabled services, where uncertainty and opacity are common, this trust-maintaining labour becomes even more vital.

Emotional labor, as I define it, is emotion management and life management combined. It is the unpaid, invisible work we do to keep those around us comfortable and happy. It envelops many other terms associated with the type of care-based labor I described in my article: emotion work, the mental load, mental burden, domestic management, clerical labor, invisible labor.

- Gemma Hartley
Author

This insight highlights certain important implications. First and foremost, trust is actively built and sustained through human interaction, and not a product of the system. Many stakeholders across the service absorb the responsibility of building trust, offering reassurance and workarounds in unpredictable scenarios. Secondly, it highlights the need for this invisible work to be acknowledged within the service design process. Doing so makes it possible to redistribute, support, or perhaps even redesign aspects of the service to reduce reliance on informal labour. There are also distinct challenges. Because invisible work is diffused and often improvised, it can be hard to observe, measure, or document. Without intentional recognition and support, this invisible labour may lead to burnout, gaps in service quality, and eroded trust over time.

This insight suggests making this trust-building work visible opens the possibility of supporting it more sustainably.

5.4 SYSTEMS MAP

The aim of this exercise was to identify points within the broader system where trust in AI-enabled public services could be supported. While my understanding of trust was already multifaceted, it remained fragmented. Developing a systems map offered a way to surface underlying dynamics and reveal connections that were previously hidden (Barbrook-Johnson & Penn, 2022; Meadows, 1999; Monat & Gannon, 2015; Morelli & Tollestrup, 2006). The results from this exercise could inform more targeted and effective design interventions.

Systems maps are powerful sense-making tools, which allow for multiple layers of a problem to be held together in one visualisation. As researched by Morelli & Tollestrup (2006), systems maps are multi-scale, multi-actor representations, that communicate new solutions across phases of the design process. In this case, I used the map to construct a trust landscape surrounding the AI-searchbot in the public sector.

The map was developed through an iterative, four step process. I began by listing all known actors involved in or affecting the AI-searchbot. These included both human actors (e.g. service creators, citizens, managers, legal teams) and non-human ones (e.g. the AI searchbot itself, databases, policy documents, trust frameworks). I grouped these into five clusters: Technical; Service Design and Delivery; Governance, Policy and Regulation; Individual Stakeholders; and Contextual and Supporting Environment.

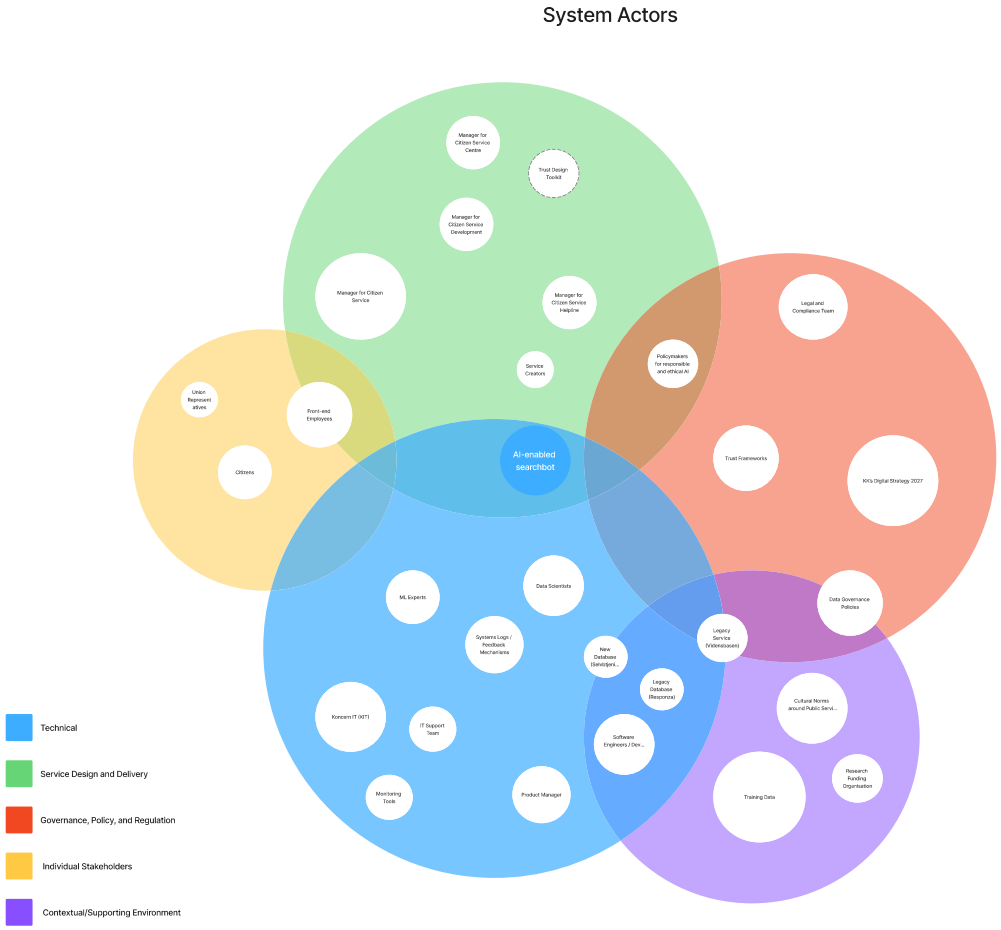


Fig. 25 Systems Map Step 1: Mapping all the actors

I then linked actors by the various interactions that occur between them. This step quickly revealed the service creators to be a central role with multiple interaction links. However, I am cautious not to overstate their centrality, as this perspective is partly shaped by my own prolonged engagement with them.

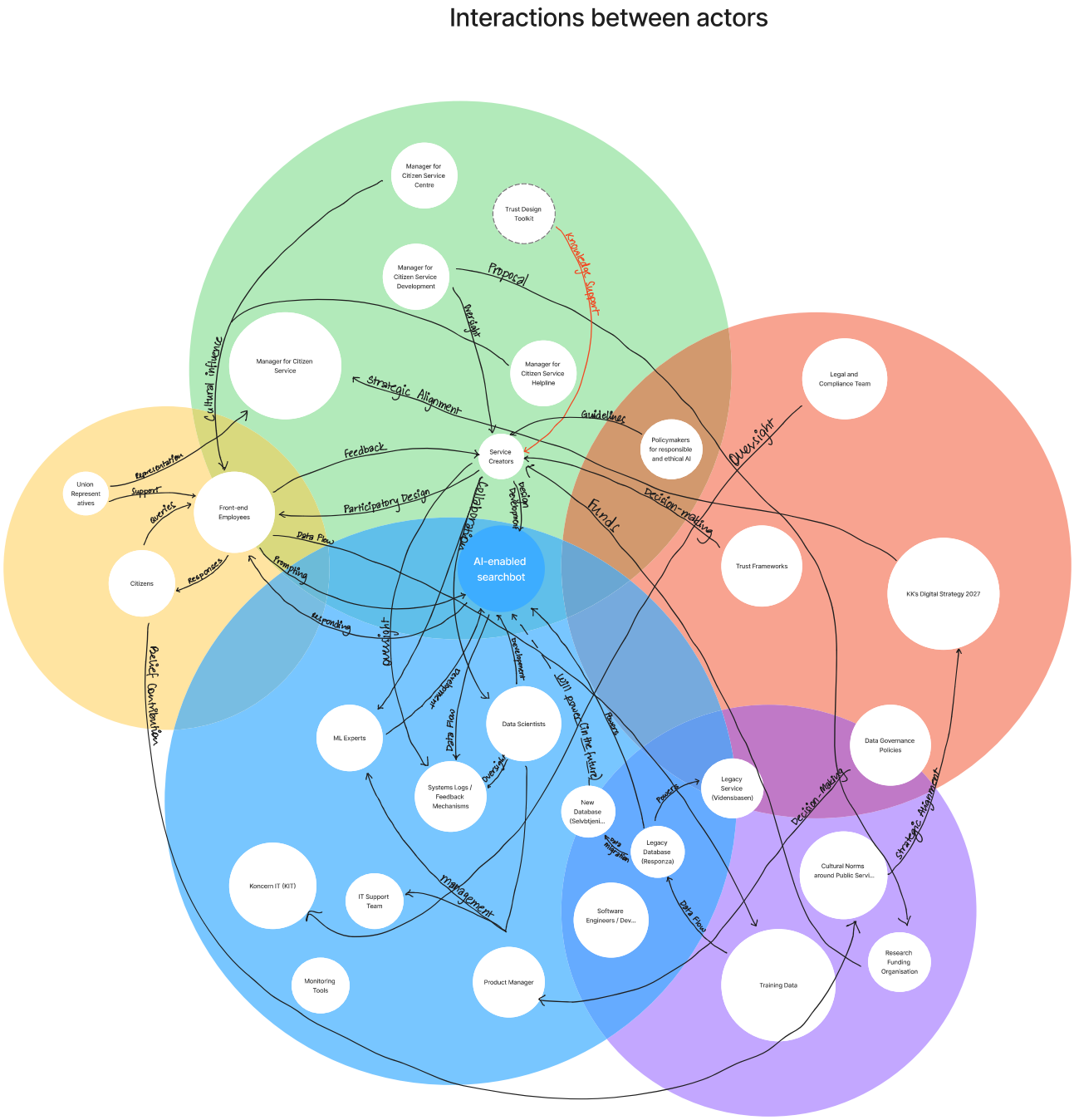


Fig. 26 Systems Map Step 2: Linking actors by interaction

I then introduced the concept of “trust touchpoints” – points in the system where trust may be formed, tested, or eroded. These touchpoints were identified and plotted on the map with a symbol underlining the situational nature of the trust. To illustrate the thought process with an example: when front-end employees get a chance to give their feedback to CSD service creators about the AI-searchbot, they feel involved in the service development process. This results in a trust-building moment for the two actors. Correspondibgly, when the feedback of front-end employees is not considered, it may lead to trust erosion.

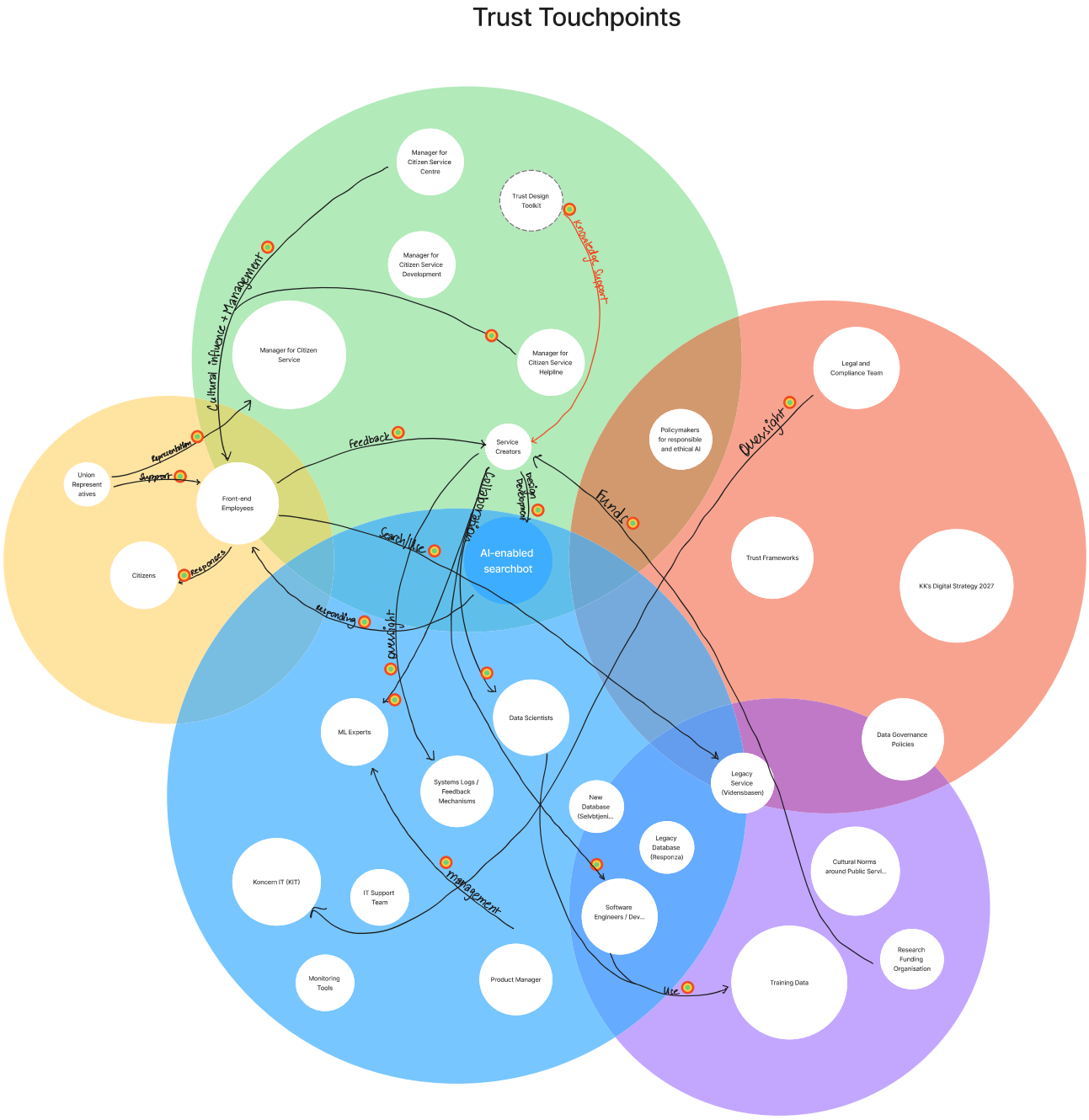
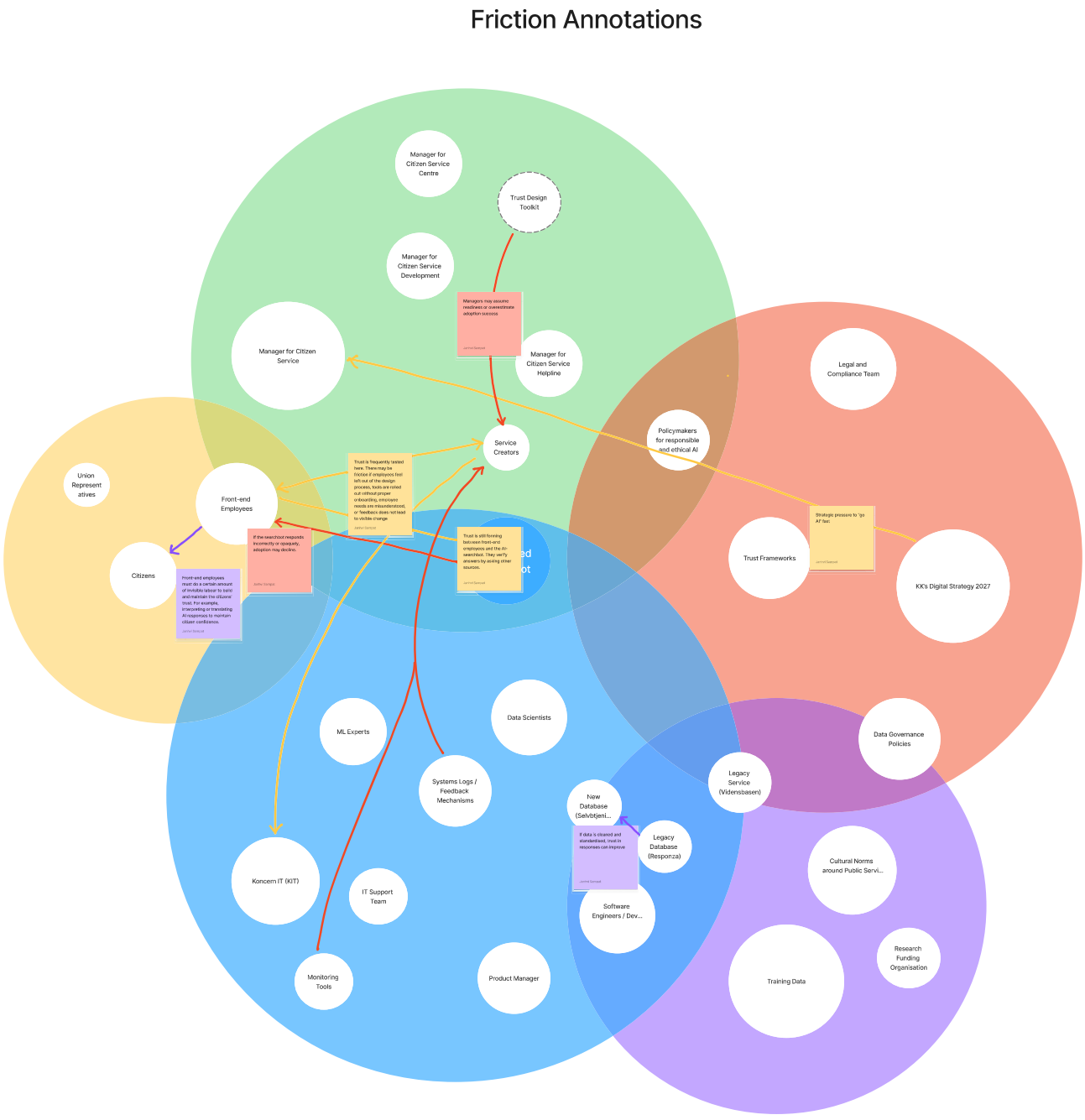


Fig. 27 Systems Map Step 3: Mapping trust touchpoints

Lastly, I annotated areas of tension in the system. Red highlighted breakdowns in trust, yellow indicated uncertainty or misalignment, and purple marked invisible work. These helped me spot pain points and underutilised gaps in the system.



🔴 Red = Trust breakdown
🟡 Yellow = Misalignment or uncertainty
🔵 Blue = Hidden opportunity or invisible work

Fig. 28 Systems Map Step 4: Annotating areas of trust

This exercise did not yield one single design direction, but it provided a richer understanding of the landscape I was working in. The mapping process itself was generative and several insights emerged. First, there was a conceptual shift in my understanding of trust. I came to understand that trust is not something built into the system, but rather an emergent property of the system that is maintained across a distributed network (Merali & Allen, 2011). The question in mind, therefore, shifted from “What can build trust here?” to “How can we maintain trust here?”. The map exposed a few blind spots, as well. Elements such as the data governance policies and trust frameworks, which generally remain invisible came to the forefront. In contrast, it highlighted certain overburdened actors, such as the front-end employees and the service creators at CSD. This insight supported the scope and focus of my project. While developing a more systemic solution was beyond my capabilities and expertise, my focus on a central conduit for trust (in this case, the service creators at CSD) could prove quite impactful. Supporting them could be a strategic lever for reinforcing trust across the wider service ecosystem. Finally, the mapping process equipped me with a language to describe the invisible dynamics (such as emotional labour, misalignment, and friction points) and a narrative framework to communicate the complexity of trust in public service systems.

5.5 HOW MIGHT WE

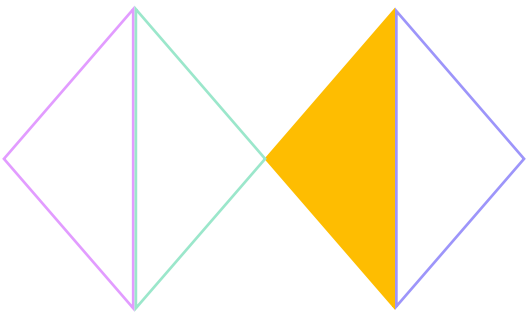
At the outset of the project, the initial research question was framed as: *"How might we design AI services in the public sector that foster trust and align with the needs of the end user?"*

Following the evolution of the project, however, the research question was reformulated as follows:

*"How might we design a **trust toolkit** to support public sector service creators in designing trustworthy AI-enabled services?"*

The revised question emphasizes the need for a framework that addresses both the end users' concerns and equips CSD service creators with tools to ensure trustworthy, transparent AI systems aligned with user needs.

5.6 DEVELOP



About This Phase

The Develop phase was about moving from insight to intervention. As a starting point, I established certain key requirements and feasibility criteria to frame my ideation process. Next, guided by the five key insights and newly refined research question, I conducted a structured ideation process to generate multiple ideas. From this ideation, a set of feasible concepts was selected, which formed the basis of the first iteration of the Trust Toolkit. The toolkit was then developed in a low-fidelity format and tested with CSD service creators.

While the toolkit itself is presented in the Deliver phase, this stage focused on its conceptual development and alignment with both research findings and user needs.

5.6.1 Feasibility and Requirements

Keeping the work culture and practices of CSD in mind, along with my capabilities and the project timeline, I defined a set of broad requirements the developed solution should fulfil:

Fig. 29 Feasibility and requirements of the Trust Toolkit

5.6.2 Ideation and Concept Development

Drawing on the insights developed in the Define phase, I carried out an explorative ideation process, in order to generate ideas on a variety of practical tools that could support the service creators at CSD in their efforts to design trustworthy AI-enabled services. This ideation process was guided by key principles from design thinking (Brown, 2008; Liedtka, 2015), which highlight the importance of exploring a broad range of possible tools, before converging on the most viable solutions. To remain grounded on usability, I kept the above feasibility and requirements in mind.

Each insight led to an actionable design prompt or How-Might-We question, after which I generated multiple ideas varying in format, complexity, and ambition using structured brainstorming. The intention was to allow for exploration of multiple tool types, such as canvases, reflection prompts, visual maps, and team activities. I then applied an internal filtering process to assess which concepts could be realistically developed and tested within the scope of this thesis. This included evaluating each concept's alignment with user needs, ease of implementation, and potential value.

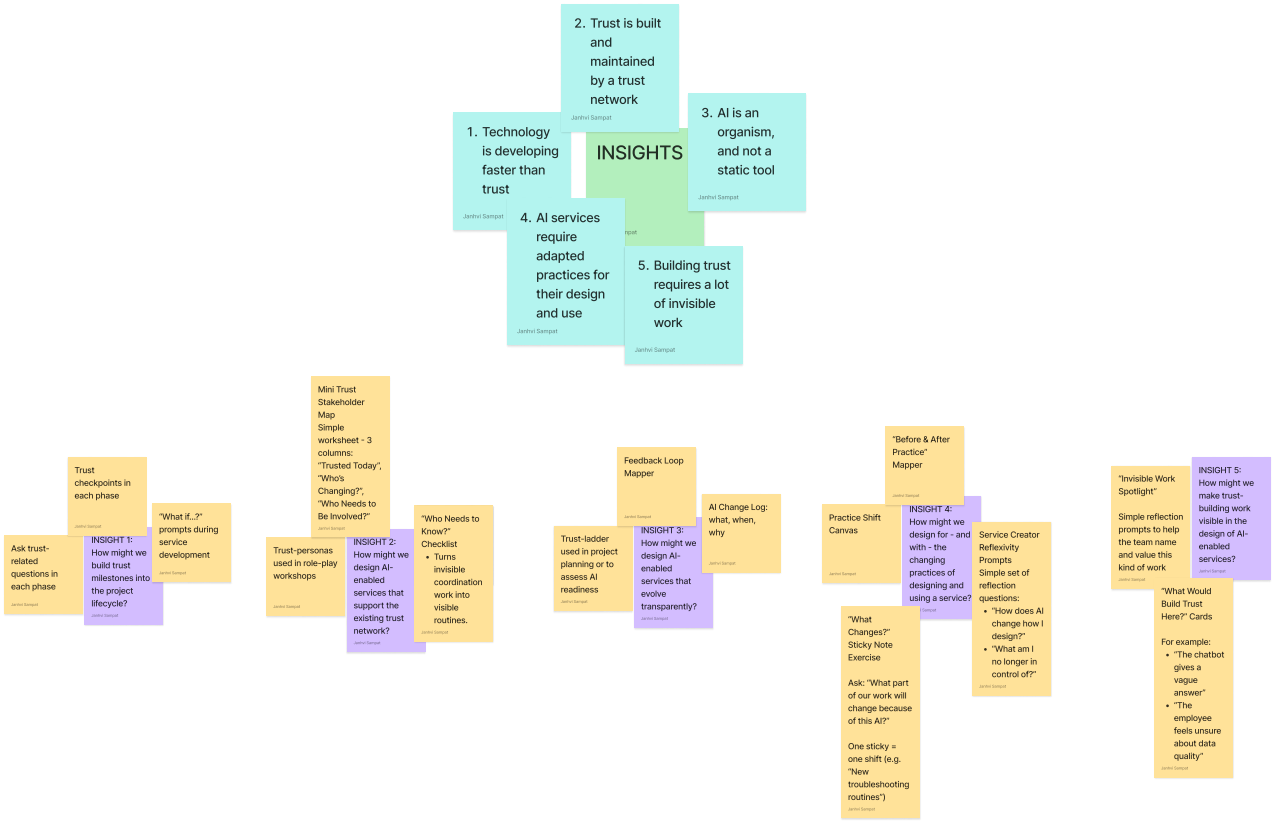


Fig. 30 Ideation and Concept Development

One idea was selected per insight as the most feasible and contextually appropriate response. This phase laid the foundation for the toolkit's prototyping and development. While many ideas were generated, the decision to proceed with a smaller set of concrete tools was deliberate, aligning with the needs and working culture of CSD's highly pragmatic and delivery-focused service creators.

5.6.3 Insight to Tool Table

To ensure the tools were grounded in the research, I carried out a structured mapping process linking each insight to a guiding How Might We question, based on the conclusion of each insight. These questions informed the direction of each tool, helping define its purpose, format, and the phase of the service creation process where it could be applied. This step ensured that the tools were not arbitrary, but directly connected to the findings. It also allowed for early reflection on how each tool might integrate into existing workflows at CSD.

The table on the following page presents this mapping. It shows the progression from insight to tool, clarifying how each design decision responds to specific trust-related challenges identified during the project.

INSIGHT	HMW	TOOL	PURPOSE	FORMAT	WHERE
Tech moves faster than trust	How might we build trust milestones into the project lifecycle?	Trust Checkpoints	Helps teams pause at key points to consider trust-building. Keeps trust aligned with tech progress.	Simple checklist	Every phase
Trust does not exist in a vacuum	How might we design AI-enabled services that support the existing trust network?	Mini Trust Stakeholder Map	Maps who is impacted by AI changes + who needs trust attention.	Worksheet	Planning + Implementation
AI is an organism and not a static tool	How might we design AI-enabled services that evolve transparently?	Feedback Loop Mapper	Helps service creators design for trust over time, not just at launch.	Canvas/Diagram	
Designing AI services reshapes the practice	How might we design for - and with - the changing practices of designing and using a service?	Service Creator Reflexivity Prompts	Encourages creators to reflect on how AI changes their design role.	Prompt Cards	Any design phase
Invisible work to establish trust	How might we make trust-building work visible in the design of AI-enabled services?	“What Would Build Trust Here?” Cards	Makes abstract trust issues tangible and designable.	Card Deck	Planning

Insights to Tool Table

5.6.4 Tool 1: Trust Checkpoints

Building on the insight about the temporal mismatch between technological development and trust development, I was prompted to consider how trust could be more intentionally addressed throughout the service creation journey at CSD. Rather than treating trust as a passive by-product of a functioning system, I reframed it as an active design concern, that could be checked at regular intervals. This reframing led to the development of a tool called the Trust Checkpoints Map. This tool integrates lightweight, reflective prompts into the existing phases of the Copenhagen Municipality's IT Project Model (Projektmateriale | It-projektrådet, n.d.).

The IT Project Model is a framework that is used across Copenhagen Municipality to guide service and IT-related projects. The model comprises six phases: Maturing; Needs Assessment; Analysis and Planning; Implementation; Completion; and Operation and Realisation (Projektmateriale | It-projektrådet, n.d.). The Trust Checkpoints Map does not aim to critique or replace this model, but to complement it, by surfacing lightweight, reflective trust-related prompts throughout these existing phases. These prompts can encourage CSD service creators to pause at key moments and consider how trust is being shaped, maintained, or challenged as the service creation journey progresses. The result was a tool, that inspiration from swimlane diagrams. The map visually organises three key elements:

• **Swimlane 1:** The official steps of the IT Project Model.
• **Swimlane 2:** Key moments in the service development journey, identified through eight months of ethnographic fieldwork with the CSD team.
• **Swimlane 3:** Targeted trust-building prompts, phrased as short reflective questions for service creators.

This mapping format is intended to be both visual and practical. This tool can be used at any point during the project lifecycle.

This mapping tool operates as a soft intervention, resonating with principles from soft systems methodology (Checkland, 1989). It is not a prescriptive tool, but rather a supportive one. By subtly shifting the mindset of service creation towards trust-building, it can help ensure that the development of trust keeps pace with the speed of technological change.

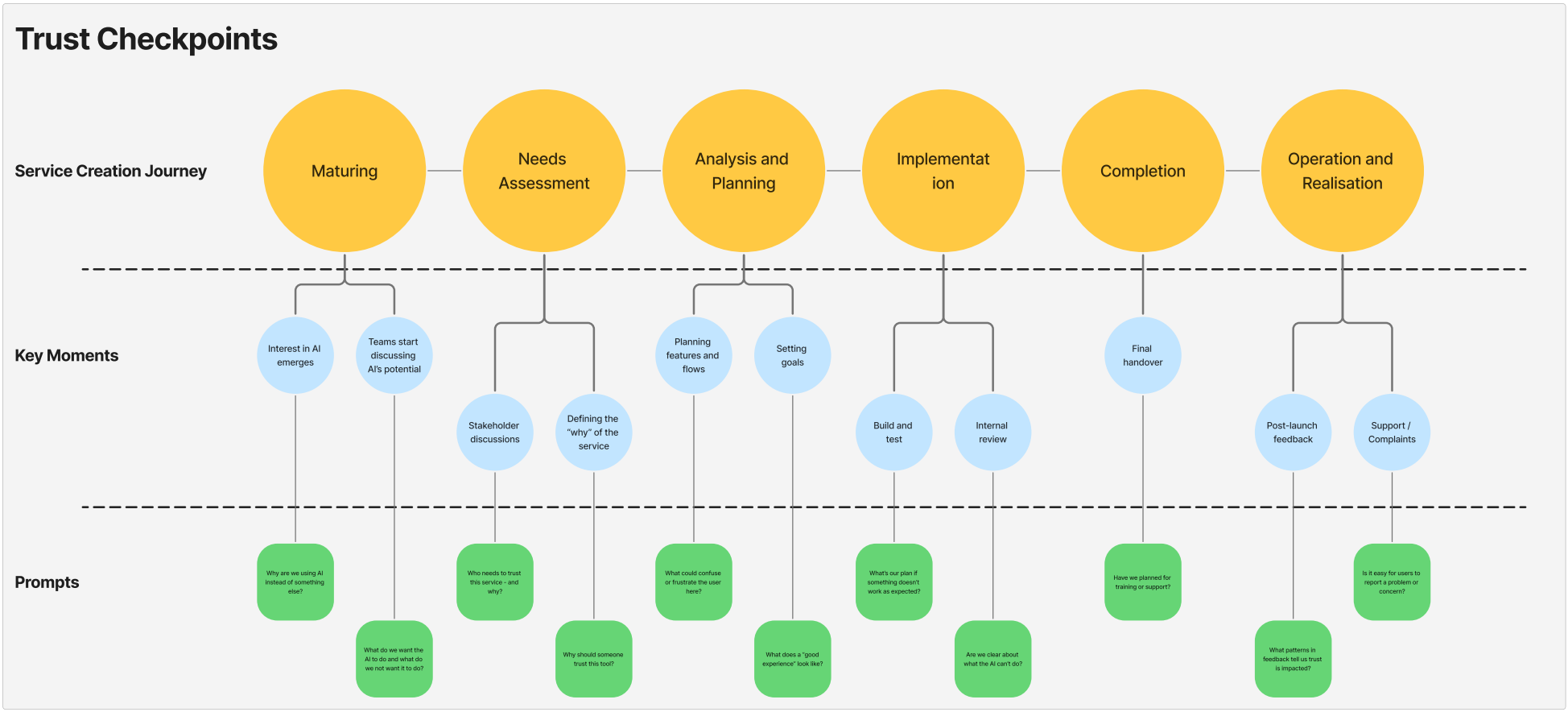


Fig. 31 Trust Checkpoints

5.6.5 Tool 2: Reflexivity Prompts

In response to the insight that designing and using AI-enabled tools actively reshapes the practice of both designing and using the service, I developed the Reflexivity Prompts tool. This set of questions is designed to help service creators at CSD engage in ongoing critical reflection on their own roles, assumptions, and the broader implications of their work. The aim is to surface and question the shifting dynamics introduced by AI solutions, particularly as they relate to responsibility, assumptions, and adaptation within public services. The prompts aim to create space for critical self-examination and team dialogue during the design process. They encourage service creators to step back and consider how AI changes their responsibilities, shapes user behaviour, and reveals implicit assumptions.

Each prompt targets a different dimension of reflexivity:

1. **"How are our roles and responsibilities changing with this AI service?"**: Encourages awareness of new duties related to ethics, data governance, and managing uncertainty.
2. **"What assumptions are we making about how the AI service will be used?"**: Surfaces potential mismatches between intended and actual use.
3. **"How will this service change what end users do and how do we support the shift?"**: Promotes anticipatory thinking around user adaptation and necessary support structures.

These prompts are designed to be used flexibly at any stage of the project. Their purpose is not to prescribe solutions but to open up space for reflective practice. This is particularly important in AI-enabled service design, where traditional design expertise is increasingly supplemented by responsibilities that are ethical, technical, and anticipatory in nature (Hess et al., 2021). Each prompt was crafted to encourage not just reflection on past actions, but a deeper reflexive awareness of how design decisions are shaped by and, in turn, shape organisational practices and user experiences. While the prompts were not formally tested, their rationale is grounded in existing literature on reflexive design and responsible innovation. Scholars have argued that reflexivity (defined as the ongoing critical examination of one's assumptions, roles, and potential impacts) is central to designing sociotechnical systems responsibly (Pihkala & Karasti, 2016; Steen, 2013). This is particularly vital in public sector contexts where AI introduces complexity and uncertainty. Reflexive practice has also been linked to adaptability in dynamic environments, helping practitioners remain responsive to emerging challenges and shifts (Steen, 2013). As such, this tool supports a more responsive, ethical, and context-sensitive approach to AI-enabled service design. They serve as a lightweight yet powerful tool to encourage more responsible and anticipatory thinking throughout the development process.

Reflexivity Prompts

What they are

Quick, reflective questions to help you step back and think about how your own role (and the roles of others) shift as you design AI-enabled services.

How to use them

- Use them alone or in a team discussion at any point in the project.
- No special format needed - grab a sticky note, journal, or chat with a colleague.
- Treat them as regular checkpoints, not just one-time reflections.

"How are our roles and responsibilities changing with this AI service?"	Helps service creators reflect on new responsibilities (e.g. data considerations, uncertainty management, ethics) that may not exist in traditional service design.
"What assumptions are we making about how the AI service will be used?"	Helps uncover gaps between how the system is designed to work and how people might actually use or trust it.
"How will this service change what end users do and how do we support the shift?"	Makes visible the effects of AI tools on users and workflows. Encourages proactive support for changing habits, expectations, and routines.

Fig. 32 Reflexivity Prompts

5.6.6 Tool 3: “What Would Build Trust Here?” Prompts

In response to the insight that trust-building is often invisible work, I developed “What Would Build Trust Here?” prompts as a tool to help CSD service creators keep trust at the forefront during the design and rollout of AI-enabled services. This tool aims to ensure that the subtle yet crucial elements of trust, such as clarity, support, consistency, and reassurance, are consistently addressed throughout the service creation process. The prompts are designed to be simple, actionable reminders that guide service creators to make thoughtful, user-centred design decisions and reinforce trust at every stage.

The prompts were designed to serve as practical checkpoints in the service creation process, particularly during planning, piloting, and launch phases. They ensure that trust-building is consistently integrated by prompting reflection on key areas that influence user confidence. These prompts encourage service creators to think critically about the user experience, helping them to identify potential gaps and address them with actionable solutions.

While the full impact of these prompts was not tested in a live service context, they were designed based on established principles of trust-building. They have the potential to guide service creators in maintaining a focus on key elements that influence trust in AI-enabled services, such as clarity, consistency, support, and reassurance in fostering user trust (Kydd, 2000; Lockey et al., 2021). By prompting service creators to consider these aspects, the tool aims to enhance the trustworthiness of AI services throughout the development process.

What-Would-Build-Trust-Here Prompts

What they are

These four prompts help service creators keep trust top-of-mind during the design and rollout of AI-enabled services. They focus on clarity, support, consistency, and reassurance - all crucial to how users build trust over time.

How to use them

- Use these prompts at key checkpoints in your service creation process, especially during planning, piloting, and launch prep.
- Don't overthink it. If a prompt sparks a new idea, small tweak, or deeper conversation, it's working!

"Is it clear to the user what the service does and doesn't do?"

"Are we giving users a simple way to ask for help?"

"Does the service feel consistent every time it is used?"

"What would reassure someone that this service is working properly?"

Fig. 33 What Would Build Trust Here? Prompts

5.6.7 Tool 4: Trust Design Canvas

This tool was developed in response to the insight that trust does not exist in a vacuum. This insight pointed towards the need for more trust-aware design practices that support the existing trust network and ease its disruption. Guided by this insight, the initial idea was a form of stakeholder trust mapping: a lightweight canvas worksheet, that could help CSD service creators identify how stakeholders might be impacted by changes in the trust network, and what concrete actions might support them in such situations. The first prototype of the tool took the shape of a three-column canvas (see Appendix B). The left column prompted the user to pinpoint stakeholders within the service/system they were designing for by asking: Who needs to trust this? The middle column invited users to identify trust needs and tensions, while the final column asked for concrete actions to support trust. However, this iteration presented certain practical challenges. It was difficult to clearly trace which needs, tensions, or actions belonged to which stakeholder, especially when overlaps emerged. In many cases, multiple stakeholders might share similar needs or concerns, which could blur the clarity of analysis.

Based on these reflections, I sought to develop a more focused tool, which centred one stakeholder group at a time. Taking inspiration from empathy maps, I redesigned the tool to a Trust Design Canvas, which places the stakeholder (for example: a citizen, frontline worker, or upper management) at the centre, surrounded by four key reflection areas:

- **Trust Needs:** What do they need to feel safe and confident?
- **Trust Breakers:** What might cause doubt, fear, or mistrust?
- **Trust Behaviours:** What signals trust (or mistrust) in how they engage with the service?
- **Trust Actions:** What could be designed to support trust in this group?

This format encourages service creators to think about how trust is experienced by different actors. It is most useful in early planning phases or when particular user groups show signs of hesitation or low adoption.

Trust Design Canvas

Project:

Stakeholder:

Use this canvas to identify trust gaps and plan small actions to build trust as you design your AI service.

Trust Breakers

What might cause doubt, fear, or mistrust?

Trust Needs

What do they need to feel safe and confident?

Trust Behaviours

What behaviours show trust or mistrust?

Trust Actions

What could we do to build trust?

Version 1.0 - Made for testing and improving.

Fig. 34 Trust Design Canvas

5.6.8 Tool 5: Feedback Loop Mapper

One of the key insights of this project was that AI is not a static tool, but an evolving organism. The insight led to the consideration of supporting more transparent evolutions of AI-enabled services. In response to this, I developed the Feedback Loop Mapper – a simple tool that supports CSD service creators in identifying and responding to trust-related signals as new AI-enabled services are implemented. The tool is particularly relevant during the later phases of a project, such as Implementation and Operation and Realisation, where the service has moved into active use and patterns of interaction emerge.

The Feedback Loop Mapper takes the form of a circle, reinforcing the ongoing and iterative nature of both AI systems and trust. It consists of five cyclical stages, each prompting a different type of reflection and response:

- **Sense:** What behaviour or signal are we noticing? (e.g. low usage, negative feedback, user workarounds)
- **Reflect:** Is the signal related to trust, usability, or another factor? What barriers might users be encountering?
- **Interpret:** Whose trust is most impacted by this signal, and in what way?
- **Act:** What concrete step could be taken to strengthen trust in response?
- **Track:** What indicators will we monitor, and how will we assess change?

This structure encourages teams to remain attentive to how trust is experienced in real-time and to treat it as an ongoing design responsibility, rather than a one-time achievement. The loop can be revisited multiple times, enabling teams to build an understanding of how trust manifests and fluctuates over time, and how design responses can shape these dynamics.

Feedback Loop Mapper

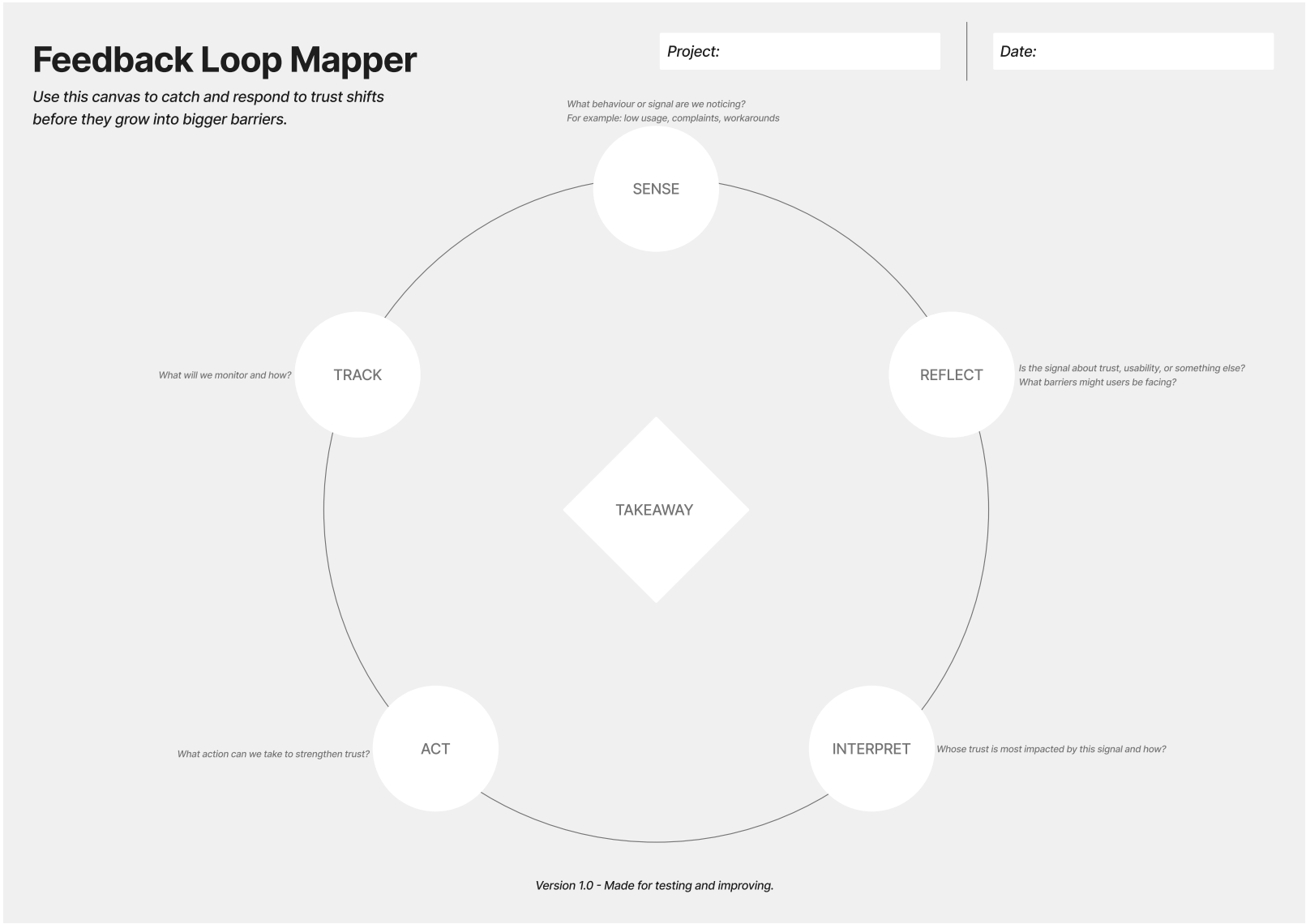


Fig. 35 Feedback Loop Mapper

5.6.9 Validating and Testing

In order to explore the usability, relevance of the toolkit, along with its practical fit within the organisational context of the service creators at CSD, I organised a validation session. The aim was to understand how the toolkit resonated with service creators involved in AI-enabled public services, and whether it supported trust-conscious design in meaningful and actionable ways.

Participants

The workshop involved four participants from CSD, all directly engaged in the development of AI-enabled services. These included: one Service Designer (involved in the AI searchbot project), one Development Consultant and Digital Project Manager (also leading the AI searchbot project), one Consultant/Project Manager (working on an AI chatbot and voicebot project), and one Digital Business Developer (also involved in the chatbot and voicebot project). Notably, the Service Designer also served as the internal supervisor of this research at CSD, and the Development Consultant and Project Manager had previously participated in research interviews for the thesis.

To ensure transparency and ethical rigour, all participants were briefed in advance and provided with a preliminary digital copy of the toolkit. Informed consent was obtained, anonymity was guaranteed, and data handling complied with GDPR (see Appendix C). All audio recordings, transcripts, and visual documentation were stored securely and deleted after the completion of the project (see Appendix C).

Workshop Format

The testing aimed to assess the clarity of each tool's purpose, its fit with existing design practices, and its potential for integration into the workflows of service creators. The session took the form of a 60-minute facilitated workshop. The background and purpose of the research were briefly presented, followed by an introduction to the toolkit and its five tools. Each tool was accompanied by a short explanation and, where relevant, an example of use. Participants were then asked to engage directly with one of the tools: the Trust Design Canvas. They were provided with printed A3 worksheets and stationery, and invited to complete the canvas based on their own project contexts. The activity was carried out in two groups of two over 20 minutes. Participants were encouraged to verbalise their thoughts during use. This was followed by a structured group discussion to elicit impressions, confusions, and suggestions for improvement.

Data Collection

The following data sources were collected during the session (see Appendix C): annotated workshop notes, audio recordings of discussions, photographic documentation, completed copies of the Trust Design Canvas.

Visual documentation of the workshop

Data Analysis

To analyse the collected data, I employed a basic form of thematic content analysis (Green & Thorogood, 2018). After the session, I revisited the audio recordings and workshop notes to extract feedback directly related to the usability, clarity, and relevance of the toolkit to the participants' workflow. Key observations and participant reflections were grouped under emerging themes, such as: clarity of instructions, design structure, contextual fit, stakeholder awareness, and perceived value of individual tools.

This inductive approach allowed for recurring patterns and concerns to surface without imposing pre-defined categories. The goal was not to perform a deep interpretative analysis, but to gather actionable insights from practitioners in order to refine the toolkit based on its real-world applicability.

Participant Reflections

The workshop surfaced several valuable reflections and suggestions:

• **Clarity and structure:** Participants noted that the Trust Design Canvas required users to reflect on three areas before identifying actionable steps. They recommended redesigning the layout to more clearly reflect this sequence.

• **Visual hierarchy:** A space to write the stakeholder group could be placed prominently in the centre of the canvas to maintain user focus.

• **Tool context:** Participants suggested including a clearer indication of when each tool is most useful in the design process.

• **Broader perspectives:** The tools helped highlight stakeholder groups often overlooked in service design. This prompted a suggestion to include speculative personas (e.g. future generations or environmental entities) as part of the toolkit's resources.

• **Meta-level discussion:** The workshop also stimulated a broader conversation about

the environmental impacts of AI-enabled services. This reflection supported the subtle prompts found in the Trust Checkpoints Map, though participants suggested this aspect could be made more explicit and impactful.

• **Institutional trust context:** The participants recognised the unusually high levels of trust in public institutions in Denmark and stressed the importance of aligning service design with citizen expectations.

• **Toolkit coherence:** The three main tools (Trust Checkpoints, Trust Design Canvas, and Feedback Loop Mapper) were seen as immediately relevant. However, there was some uncertainty around the integration of the reflective questions. One idea was to integrate them into interviews or focus group guides to better align with existing workflows.

• **Integration vs. separation:** Participants discussed whether the toolkit should be tightly integrated into existing workflows or offered as a standalone resource. Each approach had perceived benefits and drawbacks, suggesting a need for flexible implementation strategies.

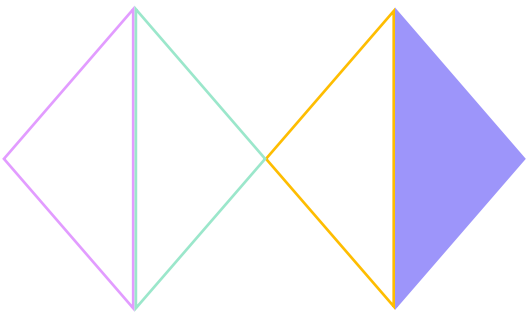
Limitations

The workshop had several constraints. The small number of participants and limited duration meant only one round of testing could be conducted. Only a single tool was tested in-depth, and none were applied in the context of an ongoing real-world project. As such, further testing will be needed to evaluate the toolkit's longitudinal impact and effectiveness in practice.

5.6.10 Next Steps

While the toolkit has reached a functional stage of development, feedback from the validation workshop highlighted areas for refinement prior to broader dissemination. These suggestions centred around improving usability, clarity, and integration into existing workflows. First, the Trust Design Canvas could benefit from small design adjustments. Participants suggested visually guiding users through the logic of the canvas i.e. filling out trust needs, trust breakers, and trust behaviours before defining trust actions. Additionally, placing the stakeholder name more prominently at the centre could help maintain focus during the exercise. Clearer indication of the phase each tool is intended to be used in was also recommended. Secondly, feedback indicated potential in providing additional resources to support broader or underrepresented perspectives, such as introducing personas (e.g. future generations or non-human actors) to prompt more expansive trust considerations. Next, participants noted a slight disconnection between the reflective questions and the tools. One possible improvement could be integrating parts of the reflective tools with the design tools, or providing guidance for using the questions during interviews or workshops. Finally, while the Trust Design Canvas was tested in the validation session, each of the other tools within the toolkit (Trust Checkpoints Map, Reflexivity Prompts, What-Would-Build-Trust-Here prompts, and the Feedback Loop Mapper) also require similar testing with end-users to evaluate their clarity, usability, and relevance. These follow-up testing sessions would be essential to ensure the entire toolkit is robust, adaptable, and well-aligned with the service creators' practices before broader implementation. While these refinements were beyond the scope of the current project, they represent clear directions for improvement should the toolkit proceed to a subsequent iteration.

5.7 DELIVER



The final deliverable of this thesis is a practical Trust Toolkit, developed to support service creators at CSD in designing for, evaluating, and maintaining trust in AI-enabled services. The toolkit is the result of insights generated during the Discover and Define phases and was shaped and validated through iterative design and feedback in the Develop phase of this project.

The toolkit is structured in two parts to reflect different moments in the service design process. The first part invites the user to “Reflect on Trust” and offers three tools to support this: Trust Checkpoints Map, Reflexivity prompts, and What Would Build Trust Here? prompts. The second part invites the user to “Design for Trust” by presenting two tools that help teams act on insights and design tangible interventions. These are the Trust Design Canvas and Feedback Loop Mapper.

The toolkit exists in a digital format and has been designed for ease of printing to support practical use in workshops, meetings, and interviews. It has been tested in part during a validation workshop with CSD service creators, where the Trust Design Canvas was testing using a real project context. Feedback gathered during this session has informed the next steps for refinement (see Section 5.6.10).

The toolkit is available as a printable PDF and is submitted as a separate document (see Appendix A). A presentation of the full toolkit, including proposed refinements, is planned for early June with key stakeholders at CSD. While developed for CSD, the toolkit has potential relevance beyond this specific context. As public sector organisations increasingly adopt AI-enabled services, tools that support value-based design and trust-aware development can play a wider role (Bak, n.d.). The toolkit may thus offer value to similar organisations exploring responsible and transparent AI deployment.

6

DISCUSSION

6.1 Addressing the Research Question

Early research revealed a conceptual and practical gap in public sector AI applications: while trust in AI is extensively discussed in ethics and governance literature (Birkstedt et al., 2023; Dignum, 2019; Morley et al., 2021), concrete tools to support trust-building at the level of everyday service design remain scarce. Public sector service designers are often left with abstract principles or technical guidelines, which do not translate easily into design decisions or workflows (EU AI Act, 2023; OECD Legal Instruments, n.d.; The Danish National Strategy for Artificial Intelligence, n.d.; Jozak, 2025). This project contributes to the bridging of this gap by introducing a practical solution that makes designing for trust more explicit, structured, and contextualised.

This thesis set out to answer the question: *"How might we design AI services in the public sector that foster trust and align with the needs of the end user?"* Initial fieldwork and literature review uncovered several key insights within trust-building for AI-enabled services. Amongst them were: the mismatch between the development of pace trust and the development of new technological innovation and a lack of visibility around trust-building work, which is often informal, emotional, and relational; and the shifting roles and practices of the stakeholders involved in AI-enabled services. These insights led to the refinement of the research question: "How might we design a trust toolkit to support public sector service creators in designing trustworthy AI-enabled services?" Through dialogue with service creators working on AI-enabled services and ethnographic and desk research, the outcome of the design process was a digital Trust Toolkit comprising five tools. These tools were iteratively developed to support the target users – the service creators within CSD. The purpose of the toolkit is to make trust building more actionable by: (i) surfacing the trust needs, risks, and expectations of stakeholders; (ii) encouraging service creators to reflect on the changing roles and practices AI-enabled services give rise to; and (iii) supporting iteration and the evaluation of trust throughout the service delivery. The toolkit was validated and partially tested by the service creators. Overall, they reacted positively to the toolkit, albeit suggesting certain design and conceptual improvements. This work builds on and extends frameworks such as Systemic Design and Value Sensitive Design, while also contributing to the Responsible Innovation (Friedman, 1996; Owen et al., 2013; Stilgoe et al., 2020).

While the project achieved its central aim, several limitations constrained the depth and scope of the work. First, the timeframe of the study constrained deeper exploration of trust as a socio-technical concept. The abstract nature of trust made scoping

and synthesis a challenge, especially in a practical organisational context. This was evident in the repeated refinement of the research question, which had to balance academic depth, organisational relevance, and personal interest. Several attempts were made to integrate a trust-based approach in early pilot rollouts of the searchbot in two departments. These were met with scepticism and were not well-received (images of these attempts can be found in Appendix B). Translating abstract, theoretical findings into actionable insights also proved to be complex. Numerous synthesis attempts were made to translate patterns, behaviours, and practices into insights, with each attempt revealing a new facade of the concept of trust. This experience suggests that even narrow angles on a high-level concept like trust can rapidly expand on closer examination. Meanwhile, CSD is a fast-paced environment where service design often prioritises optimisation and operational efficiency. While there is evident concern for responsible and user-centred design, these values are shaped by institutional priorities and invisible power structures, which influence how and where design attention is placed. Second, the time-constrained nature of the project restricted both the depth of stakeholder engagement and the extent of iterative development. Greater stakeholder involvement could have strengthened the final outcome. Third, while the toolkit was tested with some users, it has not yet been applied in a full service design cycle, meaning its long-term efficacy and adaptability remain unproven. There is also the challenge of measurement: while the toolkit aims to support trust-building, determining its impact on actual trust perceptions requires longitudinal evaluation, which was beyond the scope of this project. Future evaluation should combine qualitative methods (such as post-implementation reflection sessions and interviews) with proxy indicators (such as improved decision transparency, reduced complaints, or increased stakeholder engagement). These limitations reflect the broader challenges of applying systemic methods in time-bound, resource-limited public sector settings (Blomkamp, 2022; Leadbeater, 2013; Seddon, 2008; Virtanen & Kaivo-oja, 2015).

Nonetheless, the project demonstrates a promising direction for translating abstract values into practical design support tools. By focusing on the invisible, relational work of trust, and situating it within the reality of public service, it offers a new lens for aligning AI development with user and societal needs.

6.2 Scaling the Toolkit

As outlined in section 5.6.10, the immediate next steps for this study would be to refine the tools within the Trust Toolkit, in response to the concrete expectations and needs of the CSD service creators. These include improving the toolkit's clarity and integration into existing workflows. Addressing these points will increase the toolkit's usability and relevance in real service development contexts.

The next phase of work should focus on the real-world implementation and scaling (M.-L. Moore et al., 2015a) of the Trust Toolkit. Pilot applications in different departments within Copenhagen Municipality or other municipalities would allow assessment of the toolkit's scalability (scaling out) and adaptability across diverse service contexts. Supporting materials such as facilitation guides, templates, and training resources could be developed to make the toolkit independently usable by teams without expert support. Longitudinal studies could also track how trust-related practices and perceptions evolve over time as the toolkit is used in different contexts. In parallel, scaling up the toolkit would involve integrating the toolkit into institutional frameworks, such as internal training programmes and design or ethics guidelines. This would support its strategic influence on how AI-enabled services are evaluated and delivered. Finally, scaling deep remains critical. This involves shifting values, mindsets, and relationships within the institutions that design and deliver AI-enabled public services. The Trust Toolkit opens up space for reflection and value alignment, but more work is needed to understand how such tools can support deeper cultural transformation, particularly in contexts where efficiency is prioritised over deliberation.

Sustaining the toolkit beyond this thesis will require institutional support. Future work should clarify ownership and responsibility for maintaining, updating, and adapting the toolkit to remain relevant as technologies, regulations, and public expectations evolve. Options may include integrating it into existing service design teams, creating a stewarding role within digital governance structures, or releasing the toolkit under an open licence for broader adaptation (M.-L. Moore et al., 2015b). Beyond practical application, the Trust Toolkit creates opportunities for further academic research. Future studies could explore how tools like this facilitate value translation in design processes, particularly within complex sociotechnical systems. Comparative research across public organisations could examine how institutional context affects trust-building strategies. There is also scope to contribute to theory-building in areas such as responsible innovation, systemic design, and design for values.

6.3 Reflections on the Approach

The project applied a hybrid methodological approach that combined systemic design thinking with the Double Diamond framework. This combination allowed for a holistic understanding of trust in dynamic relationships, institutional settings, and service interactions. However, it also revealed key tensions. Systemic design is inherently non-linear and exploratory, while the Double Diamond promotes a staged, linear process. Integrating these frameworks proved complex. Deeper systemic understanding often delayed progression to design decisions. The analysis phases, while necessary, risked overwhelming the focus of the project. The approach also introduced challenges in terms of scope: the broad view encouraged by systemic thinking risked expansion

beyond manageable bounds, particularly under time constraints. Although the Design Council's existing Systemic Design framework offered some parallels, such as "orientation" and "continuing the journey", its primary focus on sustainability crises made it less suitable for the AI and trust context. Consequently, a more tailored methodological blend was pursued, allowing for context-specific flexibility.

6.4 Reflections on my Role as Designer

It is important to establish the benefit of collaborating with a public organisation for this study. As a research collaborator and employee at the Department of Citizen Service Development in Copenhagen Municipality, I had easy and immediate access to internal dynamics, tacit knowledge, and informal conversations that shaped the direction of this work. This was fundamental to the gathering of empirical data and ethnographic research, and made observing service operations and recruiting users and stakeholders relatively easy. It is also important to highlight the regular mentorship I received from both my supervisor at Aalborg University and at CSD, which led to many of the key learnings in this project. Additionally, an organised and structured approach resulted in regular incremental progress, which kept the project moving forward.

My role was to translate abstract understandings of trust into practical design tools. My close collaboration with the organisation meant that I was not a neutral observer in this environment. I was both researching and shaping the outcomes through my design choices and relationships with stakeholders. There was a constant interpretation and synthesis of theoretical frameworks, organisational constraints, and user needs. The work I did was largely relational. I facilitated conversations, reframed problems, and introduced new concepts and methods into ongoing projects. This kind of design work does not always produce immediate results, but it helps create conditions for change. In that sense, my role aligned with Manzini's (2015) view that designers enable new behaviours not by directing them, but by shaping the environment in which decisions are made.

The project shows that designers in public sector innovation roles often work as translators, facilitators, and brokers. Their impact is not just in delivering solutions, but in shifting how problems are understood and discussed. The toolkit reflects my own framing of trust as something that can be surfaced, discussed, and supported through structured reflection, and it is not a neutral product. It is a reflection of a specific designer's perspective, shaped by a particular context and set of relationships.

7

CONCLUSION

7.1 Conclusion

This research aimed to investigate how AI services in the public sector could be designed to foster trust and align with the needs of the end user. Based on the rising integration of AI-enabled services in the public sector and the crucial role of trust in the adoption of these services, the research responded to a pressing design challenge.

Applying a systemic design approach to the Double Diamond Framework, the research aimed to translate the abstract notion of "trust" into tangible tools and strategies for public service innovation. The research focused on the specific context of the Danish public sector, which reflects a high level of trust in institutions, and where the introduction of AI-enabled public services could introduce new challenges and uncertainties.

The research process highlighted that trust in AI is a multidimensional, evolving, and context-dependent phenomenon. It is shaped by cognitive, socio-technical, and cultural factors, and cannot be understood as a fixed attribute. Rather, it emerges dynamically within specific systems and relationships. This synthesis, together with insights from the broader design process, revealed a key gap: a lack of practical design tools to support public sector teams in addressing trust-related challenges when developing AI-enabled services. The outcome of this investigation was the Trust Toolkit: a design artefact developed to support public sector service creators in critically and practically engaging with the concept of trust in AI-enabled services. Developed in collaboration with Copenhagen Municipality's Department for Citizen Service Development, the toolkit was iteratively shaped through research, co-creation, and prototyping. It is designed to prompt reflection, guide decision-making, and provide structured support across various stages of service development.

While the toolkit was positively received by public sector practitioners during early testing, this study acknowledges limitations in the scale and scope of its validation. The toolkit has not yet been implemented in a full service development process, and its long-term impact on trust, adoption, and citizen outcomes remains to be studied. Additionally, the toolkit was developed within a specific cultural and institutional setting, and further research is needed to explore its adaptability in other governance contexts.

7.2 Future Research

This research resulted in the development of a practical design artefact in the form of a Trust Toolkit which is designed to support public sector service creators navigating trust-related concerns when designing AI-enabled services. While the toolkit provides one possible intervention, its format and approach should not be seen as definitive. Future research is needed to critically examine both the assumptions behind the toolkit and its actual efficacy in diverse public sector contexts. Future research should critically assess whether a toolkit is the most appropriate or effective format for addressing trust in AI, or whether alternative interventions, such as design frameworks, organisational policies, or training initiatives, might better support practice. This study was limited in scope, particularly in terms of the number of stakeholders involved and the extent of testing. Future work should evaluate the toolkit's applicability and impact across a broader range of public service contexts and user groups, ideally through longitudinal studies or implementation in live projects. Such studies could also help determine which aspects of the toolkit are most effective, and where adjustments are needed. Moreover, the research was situated in Denmark, a context marked by high institutional trust and digital maturity. Replicating the study in settings with different socio-political dynamics could surface new challenges and further test the transferability of the toolkit and its underlying assumptions. Finally, the conceptual framing of trust in this study could be re-examined or expanded. Incorporating perspectives from ethics, governance, or critical design could offer new insights into how trust is built, sustained, or eroded in AI-enabled public services. As AI technologies continue to evolve, so too must the design approaches and critical frameworks we use to engage with them.

List of References

1. About MitID - MitID. (n.d.). Retrieved 23 November 2024, from https://www.mitid.dk/en-gb/about-mitid/
2. Adams, W. (2015). Conducting Semi-Structured Interviews. https://doi.org/10.1002/9781119171386.ch19
3. Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: Progress, challenges, and future directions. Humanities and Social Sciences Communications, 11(1), 1568. https://doi.org/10.1057/s41599-024-04044-8
4. Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211.
5. An AI nation: Harnessing the opportunity of artificial intelligence in Denmark. (n.d.).
6. Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20(3), 973–989.
7. Auernhammer, J. (2020, August 11). Human-centered AI: The role of Human-centered Design Research in the development of AI. DRS2020: Synergy. https://doi.org/10.21606/drs.2020.282
8. Autopilot and Full Self-Driving (Supervised) | Tesla Support. (n.d.). Tesla. Retrieved 10 May 2025, from https://www.tesla.com/support/autopilot
9. Avin, S., Belfield, H., Brundage, M., Krueger, G., Wang, J., Weller, A., Anderljung, M., Krawczuk, I., Krueger, D., Lebensold, J., Maharaj, T., & Zilberman, N. (2021). Filling gaps in trustworthy development of AI. Science, 374(6573), 1327–1329. https://doi.org/10.1126/science.abi7176
10. Bak, A. H. (n.d.). AI i kommunerne. Retrieved 19 May 2025, from https://videncenter.kl.dk/viden-og-vaerktoejer/ai
11. Ball, J. (2004). The Double Diamond—Design Council. https://www.designcouncil.org.uk/our-resources/the-double-diamond/
12. Barbrook-Johnson, P., & Penn, A. S. (2022). Systems Mapping: How to build and use causal models of systems. Springer International Publishing. https://doi.org/10.1007/978-3-031-01919-7
13. Battistoni, C., Giraldo Nohra, C., & Barbero, S. (2019). A systemic design method to approach future complex scenarios and research towards sustainability: A holistic diagnosis tool. Sustainability, 11(16), 4458.
14. Beck, J. (2018, November 26). The Concept Creep of 'Emotional Labor'. The Atlantic. https://www.theatlantic.com/family/archive/2018/11/arlie-hochschild-housework-isnt-emotional-labor/576637/
15. Bedué, P., & Fritzsche, A. (2022). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. Journal of Enterprise Information Management, 35(2), 530–549.
16. Befolkningen 1. Januar—Statistikbanken. (n.d.). Retrieved 23 November 2024, from https://www.statistikbanken.dk/statbank5a/Graphics/MapAnalyser.asp?Maintable=BY3&lang=0&FOLKARTAET=FOLKETAL&BYER=10101100
17. Bennett, M. (1987). Real-time continuous AI systems. 134(4), 272–277.
18. Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). Hello, World: Artificial intelligence and its use in the public sector. OECD Working Papers on Public Governance, 36, 1–184.
19. Binns, R., Kleek, M. V., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). 'It's Reducing a Human Being to a Percentage'; Perceptions of Justice in Algorithmic Decisions. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3173574.3173951
20. Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: Themes, knowledge gaps and future agendas. Internet Research, 33(7), 133–167.
21. Blomkamp, E. (2022). Systemic design practice for participatory policymaking. Policy Design and Practice, 5(1), 12–31. https://doi.org/10.1080/25741292.2021.1887576
22. Bloomberg Terminal. (n.d.). Bloomberg Professional Services. Retrieved 10 May 2025, from https://www.bloomberg.com/professional/products/bloomberg-terminal/

23. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2021). On the Opportunities and Risks of Foundation Models (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2108.07258
24. Booth, R., & correspondent, R. B. S. affairs. (2024, June 23). DWP algorithm wrongly flags 200,000 people for possible fraud and error. The Guardian. https://www.theguardian.com/society/article/2024/jun/23/dwp-algorithm-wrongly-flags-200000-people-possible-fraud-error
25. Booth, R., & editor, R. B. U. technology. (2024, December 6). Revealed: Bias found in AI system used to detect UK benefits fraud. The Guardian. https://www.theguardian.com/society/2024/dec/06/revealed-bias-found-in-ai-system-used-to-detect-uk-benefits
26. Bostrom, N. (2016). Superintelligence: Paths, dangers, strategies. Oxford University Press.
27. Bostrom, N., & Yudkowsky, E. (2011). The Ethics of Artificial Intelligence.
28. Brown, T. (2008). Design Thinking. Harvard Business Review, 86(6), 84.
29. Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 2053951715622512. https://doi.org/10.1177/2053951715622512
30. Checkland, P. B. (1989). Soft systems methodology. Human Systems Management, 8(4), 273–289.
31. Chen, H., Gomez, C., Huang, C.-M., & Unberath, M. (2022). Explainable medical imaging AI needs human-centered design: Guidelines and evidence from a systematic review. Npj Digital Medicine, 5(1), 156. https://doi.org/10.1038/s41746-022-00699-2
32. Choung, H., David, P., & Ross, A. (2023). Trust in AI and Its Role in the Acceptance of AI Technologies. International Journal of Human–Computer Interaction, 39(9), 1727–1739. https://doi.org/10.1080/10447318.2022.2050543
33. Christensen, C. M. (1997). The innovator's dilemma: When new technologies cause great firms to fail (Rev., updated, and with a new chapter, [Nachdr.]). Harvard Business School Press.
34. Chun, A. C. W. (2008). Using AI for e-Government Automatic Assessment of Immigration Application Forms.
35. Coeckelbergh, M. (2012). Can we trust robots? Ethics and Information Technology, 14, 53–60.
36. Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. HBR'S 10 MUST, 67.
37. Davis, B., Glenski, M., Sealy, W., & Arendt, D. (2020). Measure utility, gain trust: Practical advice for XAI researchers. 1–8.
38. De Fine Licht, K., & De Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. AI & SOCIETY, 35(4), 917–926. https://doi.org/10.1007/s00146-020-00960-w
39. De Sousa, W. G., de Melo, E. R. P., 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.
40. Definition of TRUST. (2025, May 14). https://www.merriam-webster.com/dictionary/trust
41. Delgado, F., Yang, S., Madaio, M., & Yang, Q. (2023). The participatory turn in ai design: Theoretical foundations and the current state of practice. 1–23.
42. Denmark Tops Europe in AI Adoption. (2025, March 31). Invest In Denmark. https://idk.editor.um.dk/insights/denmark-tops-europe-in-ai-adoption-a-testbed-for-tomorrows-ai-breakthroughs
43. Digital Fuldmagt | Københavns Kommune. (n.d.). Retrieved 2 May 2025, from https://www.kk.dk/borger/borgerservice/mitid-og-digitale-tjenester/digital-fuldmagt
44. Digital Post | Københavns Kommune. (n.d.). Retrieved 2 May 2025, from https://www.kk.dk/borger/borgerservice/mitid-og-digitale-tjenester/digital-post
45. Digitaliseringsstrategi 2024-2027 – Et bedre københavnerliv. (n.d.).

46. Dignum, V. (2019). Responsible artificial intelligence: How to develop and use AI in a responsible way (Vol. 2156). Springer.
47. Dove, G., Halskov, K., Forlizzi, J., & Zimmerman, J. (2017). UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 278–288. https://doi.org/10.1145/3025453.3025739
48. Driankov, D., Hellendoorn, H., & Reinfrank, M. (1996). An Introduction to Fuzzy Control. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-03284-8
49. Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda (An opinion paper for International Journal of Information Management).
50. Eggers, W. D., Schatsky, D., & Viechnicki, P. (2017, April 26). AI-augmented government. Deloitte Insights. https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/artificial-intelligence-government.html
51. Employee.kk.dk | Employee in Copenhagen Municipality. (n.d.). Retrieved 23 November 2024, from https://medarbejder.kk.dk/
52. EU AI Act: First regulation on artificial intelligence. (2023, August 6). Topics | European Parliament. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
53. Eubanks, V. (2019). Automating inequality: How high-tech tools profile, police, and punish the poor (First Picador edition). Picador St. Martin's Press.
54. Evans, M. M., & Wensley, A. K. (2009). Predicting the influence of network structure on trust in knowledge communities: Addressing the interconnectedness of four network principles and trust. Electronic Journal of Knowledge Management, 7(1), pp41-54.
55. Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. Sci, 6(1), 3.
56. Ferrario, A., Loi, M., & Viganò, E. (2020). In AI we trust incrementally: A multi-layer model of trust to analyze human-artificial intelligence interactions. Philosophy & Technology, 33(3), 523–539.
57. Forsyth, P. B., Adams, C. M., & Hoy, W. K. (2011). Collective trust. Why Schools Can't Improve, 101–171.
58. Friedman, B. (1996). Value-sensitive design. Interactions, 3(6), 16–23.
59. Gaver, W. (2012). What should we expect from research through design? Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 937–946. https://doi.org/10.1145/2207676.2208538
60. Giarratano, J. C., & Riley, G. (2006). Expert systems: Principles and programming (4. ed., [repr.]). Thomson Course Technology.
61. Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. Annual Review of Psychology, 62(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346
62. Giordano, F., Morelli, N., Götzen, A. D., & Hunziker, J. (2018). The Stakeholder map: A conversation tool for designing people-led public services.
63. Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057
64. Goertzel, B., & Pennachin, C. (2007). Artificial general intelligence. Springer.
65. Grammarly: Free AI Writing Assistance. (n.d.). Retrieved 10 May 2025, from https://www.grammarly.com/
66. Green, J., & Thorogood, N. (2018). Qualitative methods for health research.
67. Guo, W. (2020). Explainable artificial intelligence for 6G: Improving trust between human and machine. IEEE Communications Magazine, 58(6), 39–45.
68. Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial Intelligence Trust, Risk and Security Management (AI TRiSM): Frameworks, applications, challenges and future research directions. Expert Systems with Applications, 240, 122442. https://doi.

org/10.1016/j.eswa.2023.122442

69. Hancock, P. A., Billings, D. R., & Schaefer, K. E. (2011). Can You Trust Your Robot? Ergonomics in Design: The Quarterly of Human Factors Applications, 19(3), 24–29. https://doi.org/10.1177/1064804611415045
70. Hang, H., & Chen, Z. (2022). How to realize the full potentials of artificial intelligence (AI) in digital economy? A literature review. Journal of Digital Economy, 1(3), 180–191. https://doi.org/10.1016/j.jdec.2022.11.003
71. Harari, Y. N. (2024). Al Will Take Over Human Systems From Within. https://www.noemamag.com/al-will-take-over-human-systems-from-within
72. Hasija, A., & Esper, T. L. (2022). In artificial intelligence (AI) we trust: A qualitative investigation of AI technology acceptance. Journal of Business Logistics, 43(3), 388–412.
73. Hess, D. J., Lee, D., Biebl, B., Fränzle, M., Lehnhoff, S., Neema, H., Niehaus, J., Pretschner, A., & Sztipanovits, J. (2021). A comparative, sociotechnical design perspective on Responsible Innovation: Multidisciplinary research and education on digitized energy and Automated Vehicles. Journal of Responsible Innovation, 8(3), 421–444. https://doi.org/10.1080/23299460.2021.1975377
74. Hieronymi, P. (2008). The reasons of trust. Australasian Journal of Philosophy, 86(2), 213–236.
75. Higgins, E. T., & Kruglanski, A. W. (2007). Social psychology: Handbook of basic principles (Second edition). Guilford.
76. Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. Human Factors, 57(3), 407–434. https://doi.org/10.1177/0018720814547570
77. Huang, M.-H., & Rust, R. T. (2018). Artificial Intelligence in Service. Journal of Service Research, 21(2), 155–172. https://doi.org/10.1177/1094670517752459
78. Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI (arXiv:2010.07487). arXiv. https://doi.org/10.48550/arXiv.2010.07487
79. Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: The global landscape of ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
80. Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. Journal of Business Research, 58(4), 500–507.
81. Jones, P. H. (2014). Systemic design principles for complex social systems. Social Systems and Design, 91–128.
82. Jørgensen, R. F. (2023). Data and rights in the digital welfare state: The case of Denmark. Information, Communication & Society, 26(1), 123–138.
83. Joseph, R. C. (2010). Individual resistance to IT innovations. Communications of the ACM, 53(4), 144–146. https://doi.org/10.1145/1721654.1721693
84. Jozak, T. (2025). 2024 EU AI Act: A detailed analysis. Available at SSRN 5196856.
85. Jurafsky, D., & Martin, J. H. (2000). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall.
86. Kahneman, D. (2011). Thinking, fast and slow. macmillan.
87. Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), 263–291. JSTOR. https://doi.org/10.2307/1914185
88. Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, 62(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004
89. Kesari, G. (n.d.). Next-Gen Smart Warehouses: How AI Is Shaping The Modern Supply Chain. Forbes. Retrieved 10 May 2025, from https://www.forbes.com/sites/ganeskesari/2024/06/26/smart-warehouses-how-ai-is-shaping-the-modern-supply-chain/

90. Kim, D. H. (1999). Introduction to systems thinking (Vol. 16). Pegasus Communications Waltham, MA.
91. Kim, J., Giroux, M., & Lee, J. C. (2021). When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. Psychology & Marketing, 38(7), 1140–1155. https://doi.org/10.1002/mar.21498
92. Københavns Borgerservice | Københavns Kommune. (n.d.). Retrieved 24 May 2025, from https://www.kk.dk/borger/borger-service
93. Kramer, R. M. (2010). Collective trust within organizations: Conceptual foundations and empirical insights. Corporate Reputation Review, 13, 82–97.
94. Kramer, R. M., Brewer, M. B., & Hanna, B. A. (1996). Collective trust and collective action: The decision to trust as a social decision. Earlier versions of this chapter were presented at the Asilomar Conference on Organizations and the Society of Experimental Social Psychology meetings.
95. Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2015). Accountable Algorithms. University of Pennsylvania Law Review, 165.
96. Kumar, D., Pratap, B., & Aggarwal, A. (2021). Public trust in state governments in India: Who are more confident and what makes them confident about the government? Asian Journal of Comparative Politics, 6(2), 154–174.
97. Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. Telecommunications Policy, 44(6), 101976. https://doi.org/10.1016/j.telpol.2020.101976
98. Kydd, A. (2000). Trust, Reassurance, and Cooperation. International Organization, 54(2), 325–357. https://doi.org/10.1162/002081800551190
99. Kyung, N., & Kwon, H. E. (2022). Rationally trust, but emotionally? The roles of cognitive and affective trust in laypeople's acceptance of AI for preventive care operations. Production and Operations Management. https://doi.org/10.1111/poms.13785
100. Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. Journal of the Association for Information Systems, 16(10), 1.
101. Larsen, C. A. (2013). The rise and fall of social cohesion: The construction and de-construction of social trust in the US, UK, Sweden and Denmark. Oxford University Press, USA.
102. Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford university press.
103. Lawson, B. (2010). How designers think: The design process demystified (Reprint). Elsevier Architectural Press.
104. Leadbeater, C. (2013). The systems innovator. Mulgan G. and Leadbeater C. Systems Innovation. Discussion Paper. January. London: NESTA, 25–56.
105. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
106. Lee, M. K., & Rich, K. (2021). Who is included in human perceptions of AI?: Trust and perceived fairness around healthcare AI and cultural mistrust. 1–14.
107. Levi, M., & Stoker, L. (2000). Political Trust and Trustworthiness. Annual Review of Political Science, 3(1), 475–507. https://doi.org/10.1146/annurev.polisci.3.1.475
108. Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. Journal of Management, 32(6), 991–1022.
109. Lewicki, R. J., & Wiethoff, C. (2000). Trust, trust development, and trust repair. The Handbook of Conflict Resolution: Theory and Practice, 1(1), 86–107.
110. Li, M., Yin, D., Qiu, H., & Bai, B. (2021). A systematic review of AI technology-based service encounters: Implications for hospitality and tourism operations. International Journal of Hospitality Management, 95, 102930. https://doi.org/10.1016/j.ijhm.2021.102930

111. Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. The Journal of Strategic Information Systems, 17(1), 39–71. https://doi.org/10.1016/j.jsis.2008.01.001
112. Liao, Q. V., & Sundar, S. S. (2022). Designing for Responsible Trust in AI Systems: A Communication Perspective. 1257–1268. https://doi.org/10.1145/3531146.3533182
113. Liedtka, J. (2015). Perspective: Linking Design Thinking with Innovation Outcomes through Cognitive Bias Reduction. Journal of Product Innovation Management, 32(6), 925–938. https://doi.org/10.1111/jpim.12163
114. Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2021.664
115. Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–16. https://doi.org/10.1145/3313831.3376727
116. Lorenzen, M. S. (2024). Digital Democracy Centre Undersøger: Danskernes forståelse af, holdninger til og brug af (generativ) kunstig intelligens. University of Southern Denmark. https://doi.org/10.21996/GPMZ-S343
117. Lyrio, M. V. L., Lunkes, R. J., & Taliani, E. T. C. (2018). Thirty Years of Studies on Transparency, Accountability, and Corruption in the Public Sector: The State of the Art and Opportunities for Future Research. Public Integrity, 20(5), 512–533. https://doi.org/10.1080/10999922.2017.1416537
118. Malik, Y. K. (1979). Trust, efficacy, and attitude toward democracy: A case study from India. Comparative Education Review, 23(3), 433–442.
119. Manzini, E. (2015). Design, when everybody designs: An introduction to design for social innovation. MIT press.
120. Mardiani, E., Judijanto, L., & Rukmana, A. Y. (2023). Improving trust and accountability in AI systems through technological era advancement for decision support in Indonesian manufacturing companies. West Science Interdisciplinary Studies, 1(10), 1019–1027.
121. Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. Nature, 568(7751), 163–165.
122. Markides, C. (2006). Disruptive Innovation: In Need of Better TheoryÃ.
123. Matthews, T., Judge, T., & Whittaker, S. (2012). How do designers and user experience professionals actually perceive and use personas? Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1219–1228. https://doi.org/10.1145/2207676.2208573
124. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. The Academy of Management Review, 20(3), 709. https://doi.org/10.2307/258792
125. Mazurek, G., & Małagocka, K. (2019). Perception of privacy and data protection in the context of the development of artificial intelligence. Journal of Management Analytics, 6(4), 344–364. https://doi.org/10.1080/23270012.2019.1671243
126. McTear, M. (2021). Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots. Springer International Publishing. https://doi.org/10.1007/978-3-031-02176-3
127. Meadows, D. (1999). Leverage points. Places to Intervene in a System, 19, 28.
128. Medaglia, R., Gil-Garcia, J. R., & Pardo, T. A. (2023). Artificial Intelligence in Government: Taking Stock and Moving Forward. Social Science Computer Review, 41(1), 123–140. https://doi.org/10.1177/08944393211034087
129. Medeiros, M. (2020). Public and private dimensions of AI technology and security. Modern Conflict and Artificial Intelligence, 20.
130. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys, 54(6), 1–35. https://doi.org/10.1145/3457607
131. Mensah, G. B. (2023). Artificial intelligence and ethics: A comprehensive review of bias mitigation, transparency, and account-

ability in AI Systems. Preprint, November, 10(1).

132. Merali, Y., & Allen, P. (2011). Complexity and systems thinking. The SAGE Handbook of Complexity and Management, 31–52.
133. Mergel, I., Dickinson, H., Stenvall, J., & Gasco, M. (2023). Implementing AI in the public sector. Public Management Review, 1–14. https://doi.org/10.1080/14719037.2023.2231950
134. Miaskiewicz, T., & Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes? Design Studies, 32(5), 417–430. https://doi.org/10.1016/j.destud.2011.03.003
135. Mikhaylov, S. J., Esteve, M., & Campion, A. (2018). Artificial intelligence for the public sector: Opportunities and challenges of cross-sector collaboration. Philosophical Transactions of the Royal Society a: Mathematical, Physical and Engineering Sciences, 376(2128), 20170357.
136. Miller, T. (2022). Are we measuring trust correctly in explainability, interpretability, and transparency research? arXiv Preprint arXiv:2209.00651.
137. Mitchell, T. M. (1997). Machine Learning. McGraw-Hill.
138. MitID is Denmark's digital ID - MitID. (n.d.). Retrieved 2 May 2025, from https://www.mitid.dk/en-gb/
139. Mollering, G. (2006). Trust: Reason, routine, reflexivity. Emerald Group Publishing.
140. Monat, J. P., & Gannon, T. F. (2015). Systems Thinking.
141. Moore, G. A. (1991). Crossing the Chasm.
142. Moore, M.-L., Riddell, D., & Vocisano, D. (2015a). Scaling Out, Scaling Up, Scaling Deep: Strategies of Non-profits in Advancing Systemic Social Innovation. Journal of Corporate Citizenship, 2015(58), 67–84. https://doi.org/10.9774/GLEAF.4700.2015.ju.00009
143. Moore, M.-L., Riddell, D., & Vocisano, D. (2015b). Scaling Out, Scaling Up, Scaling Deep: Strategies of Non-profits in Advancing Systemic Social Innovation. The Journal of Corporate Citizenship, 58, 67–84.
144. Morelli, N., & Tollestrup, C. (2006). NEW REPRESENTATION TECHNIQUES FOR DESIGNING IN A SYSTEMIC PERSPECTIVE.
145. Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: A pragmatic operationalisation of AI ethics. Minds and Machines, 31(2), 239–256.
146. Murad, C., & Munteanu, C. (2020). Designing voice interfaces: Back to the (curriculum) basics. 1–12.
147. Nielsen, J. H., & Lindvall, J. (2021). Trust in government in Sweden and Denmark during the COVID-19 epidemic. West European Politics, 44(5–6), 1180–1204.
148. Nissenbaum, H. (2001). SECURING TRUST ONLINE: WISDOM OR OXYMORON? BOSTON UNIVERSITY LAW REVIEW, 81.
149. OECD Legal Instruments. (n.d.). Retrieved 20 May 2025, from https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
150. OECD Survey on Drivers of Trust in Public Institutions 2024 Results - Country Notes: Denmark. (2024, July 10). OECD. https://www.oecd.org/en/publications/oecd-survey-on-drivers-of-trust-in-public-institutions-2024-results-country-notes_a8004759-en/denmark_ac5b6973-en.html
151. Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A Framework for Responsible Innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), Responsible Innovation (1st ed., pp. 27–50). Wiley. https://doi.org/10.1002/9781118551424.ch2
152. Park, E. H., Werder ,Karl, Cao ,Lan, & and Ramesh, B. (2022). Why do Family Members Reject AI in Health Care? Competing Effects of Emotions. Journal of Management Information Systems, 39(3), 765–792. https://doi.org/10.1080/07421222.2022.2096550
153. Patrício, L., Fisk, R. P., Falcão E Cunha, J., & Constantine, L. (2011). Multilevel Service Design: From Customer Value Constellation

to Service Experience Blueprinting. Journal of Service Research, 14(2), 180–200. https://doi.org/10.1177/1094670511401901

154. Pieters, W. (2011). Explanation and trust: What to tell the user in security and AI? Ethics and Information Technology, 13, 53–64.
155. Pihkala, S., & Karasti, H. (2016). Reflexive engagement: Enacting reflexivity in design and for 'participation in plural'. Proceedings of the 14th Participatory Design Conference: Full Papers - Volume 1, 21–30. https://doi.org/10.1145/2940299.2940302
156. Projektmateriale | It-projektrådet. (n.d.). Retrieved 17 May 2025, from https://itprojektraad.kk.dk/for-projekter/projektmateriale
157. Public attitudes toward new technologies: Our post-truth, post-trust, post-expert world demands a deeper understanding of the factors that drive public attitudes. (n.d.). https://doi.org/10.1177/0036850419851350
158. Rhydderch, A. (2017). Scenario Building: The 2x2 Matrix. Futuribles International.
159. Richmond, B. (1993). Systems thinking: Critical thinking skills for the 1990s and beyond. System Dynamics Review, 9(2), 113–133. https://doi.org/10.1002/sdr.4260090203
160. Riegelsberger, J., Sasse, M. A., & McCarthy, J. D. (2005). The mechanics of trust: A framework for research and design. International Journal of Human-Computer Studies, 62(3), 381–422. https://doi.org/10.1016/j.ijhcs.2005.01.001
161. Rogers, E. M. (1962). Diffusion of Innovations. Free Press of Glencoe.
162. Romzek, B. S. (2000). Dynamics of Public Sector Accountability in an Era of Reform. International Review of Administrative Sciences, 66(1), 21–44. https://doi.org/10.1177/0020852300661004
163. Rosenbaum, M. S., Otalora, M. L., & Ramírez, G. C. (2017). How to create a realistic customer journey map. Business Horizons, 60(1), 143–150. https://doi.org/10.1016/j.bushor.2016.09.010
164. Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Introduction to Special Topic Forum: Not so Different after All: A Cross-Discipline View of Trust. The Academy of Management Review, 23(3), 393–404.
165. Russell, S. J., & Norvig, P. (2016). Artificial intelligence: A modern approach. pearson.
166. Ryan, A. (2014). A Framework for Systemic Design. FormAkademisk - Forskningstidsskrift for Design Og Designdidaktikk, 7(4). https://doi.org/10.7577/formakademisk.787
167. Ryan, M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability. Science and Engineering Ethics, 26(5), 2749–2767.
168. Salminen, J., Jansen, B. J., An, J., Kwak, H., & Jung, S.-G. (2018). ARE PERSONAS DONE? EVALUATING THE USEFULNESS OF PERSONAS IN THE AGE OF ONLINE ANALYTICS. Persona Studies, 4(2).
169. Schilke, O., & Cook, K. S. (2013). A cross-level process theory of trust development in interorganizational relationships. Strategic Organization, 11(3), 281–303.
170. Schmidt, P., Biessmann, F., & Teubner, T. (2020). Transparency and trust in artificial intelligence systems. Journal of Decision Systems, 29(4), 260–278.
171. Seddon, J. (2008). Systems Thinking in the Public Sector. Triarchy Press.
172. Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies, 146, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551
173. Shroff, G. (2013). The Intelligent Web: Search, smart algorithms, and big data. OUP Oxford.
174. Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics.
175. Sønderskov, K. M., & Dinesen, P. T. (2014). Danish exceptionalism: Explaining the unique increase in social trust over the past 30 years. European Sociological Review, 30(6), 782–795.
176. Sønderskov, K. M., & Dinesen, P. T. (2016). Trusting the state, trusting each other? The effect of institutional trust on social trust. Political Behavior, 38, 179–202.
177. Star, S. L., & Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. Computer Supported

Cooperative Work (CSCW), 8, 9–30.

178. Steen, M. (2013). Co-design as a process of joint inquiry and imagination. Design Issues, 29(2), 16–28.
179. Stilgoe, J. (2023). What does it mean to trust a technology? Science, 382(6676), eadm9782. https://doi.org/10.1126/science.adm9782
180. Stilgoe, J., Owen, R., & Macnaghten, P. (2020). Developing a framework for responsible innovation. In The ethics of nanotechnology, geoengineering, and clean energy (pp. 347–359). Routledge.
181. Taddeo, M. (2009). Defining trust and e-trust: From old theories to new problems. International Journal of Technology and Human Interaction (IJTHI), 5(2), 23–35.
182. Tag et kørekort | Københavns Kommune. (n.d.). Retrieved 2 May 2025, from https://www.kk.dk/borger/borgerservice/koerekort/tag-et-koerekort
183. Thalpage, N. (2023). Unlocking the Black Box: Explainable Artificial Intelligence (XAI) for Trust and Transparency in AI Systems. Journal of Digital Art & Humanities, 4(1), 31–36. https://doi.org/10.33847/2712-8148.4.1_4
184. The Danish National Strategy for Artificial Intelligence. (n.d.). Retrieved 20 May 2025, from https://en.digst.dk/strategy/the-danish-national-strategy-for-artificial-intelligence/
185. The organization of the Culture and Leisure Administration | Copenhagen Municipality. (n.d.). Retrieved 23 November 2024, from https://www.kk.dk/om-kommunen/forvaltninger/kultur-og-fritidsforvaltningen/organisation
186. Ubaldi, B., Le Fevre, E. M., Petrucci, E., Marchionni, P., Biancalana, C., Hiltunen, N., Intravaia, D. M., & Yang, C. (2019). State of the art in the use of emerging technologies in the public sector (OECD Working Papers on Public Governance 31; OECD Working Papers on Public Governance, Vol. 31). https://doi.org/10.1787/932780bc-en
187. van Kersbergen, K., & Tinggaard Svendsen, G. (2024). Social trust and public digitalization. AI & SOCIETY, 39(3), 1201–1212.
188. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. Advances in Neural Information Processing Systems, 30. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
189. 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
190. Virtanen, P., & Kaivo-oja, J. (2015). Public service systems and emerging systemic governance challenges. International Journal of Public Leadership, 11(2), 77–91. https://doi.org/10.1108/IJPL-02-2015-0007
191. Von Eschenbach, W. J. (2021). Transparency and the Black Box Problem: Why We Do Not Trust AI. Philosophy & Technology, 34(4), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0
192. Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. Business Process Management Journal, 26(7), 1893–1924. https://doi.org/10.1108/BPMJ-10-2019-0411
193. Weller, A. (2019). Transparency: Motivations and challenges. In Explainable AI: interpreting, explaining and visualizing deep learning (pp. 23–40). Springer.
194. What is an aircraft autopilot? (2024, April 29). https://worldaviationato.com/en/aircraft-autopilot/
195. Why do people resist new technologies? History might provide the answer. (2016, July 6). World Economic Forum. https://www.weforum.org/stories/2016/07/why-do-people-resist-new-technologies-history-has-answer/
196. Wilson, C., & Van Der Velden, M. (2022). Sustainable AI: An integrated model to guide public sector decision-making. Technology in Society, 68, 101926. https://doi.org/10.1016/j.techsoc.2022.101926

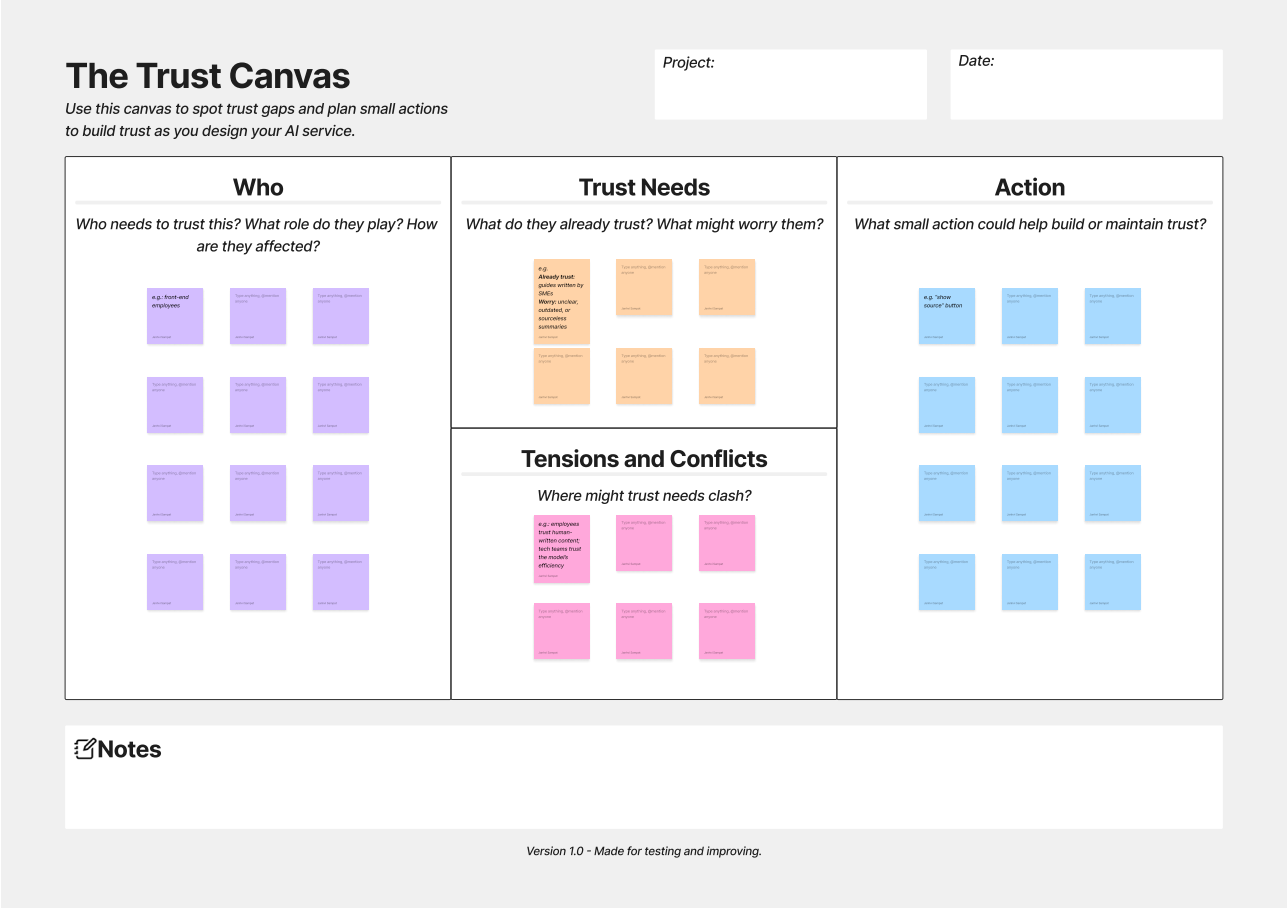
197. Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector—Applications and Challenges. International Journal of Public Administration, 42(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103
198. Wong, L.-W., Tan, G. W.-H., Ooi, K.-B., & Dwivedi, Y. (2024). The role of institutional and self in the formation of trust in artificial intelligence technologies. Internet Research, 34(2), 343–370.

Appendix A

See document: The Trust Toolkit *(PDF, 20 pages)* or click the link here.

Appendix B

Appendix B includes the documentation of the design process. Click here to view the digital workspace used during this project.



Trust Design Canvas Iteration 1

Trust-focused implementation attempts of the AI-Searchbot: (L) including the word “trustworthy” in communication, and (R) introducing a “knowledge-sharing” wall to encourage collective trust building.

Appendix C

Appendix C includes the materials used and data collected for the interviews and workshop. Transcripts and audio recordings available upon request.

Category	Research Question	Interview Question
Introduction	What is my thesis about?	I'm working on my Service Systems Design master's thesis with the Municipality of Copenhagen's Citizen Service Development department. As I mentioned on LinkedIn, the municipality is developing AI solutions to integrate into public services, both internally and externally. One of these tools is now ready for gradual rollout to front-line employees. It's a search-bot designed to replace the internal data system, helping employees quickly find answers to complex citizen queries. A key challenge with this rollout is getting end-users to trust these AI solutions enough to use them. This has led me to investigate the design principles for trustworthy AI in the public sector.
	What is the purpose of this conversation?	I'm really looking forward to hearing about your experience with AI projects like the VELUX advisor. I'm especially interested in how you built trust and reassurance into the design and how that might compare to the public sector.
	How long is this conversation?	I have about 7-10 questions regarding your experience, so I expect this conversation to last about an hour.
	Consent for recording and handling of data	
		Can you tell me more about your work on the VELUX AI advisor project? What were its main goals and challenges?
		What role did trust and transparency play in your AI advisor project? Were there specific features or design decisions aimed at building these?
		You mentioned that reassurance was a key finding in your research—could you elaborate on how this was identified as important and how it was addressed in the AI design?
		How did you approach understanding user expectations around reassurance? Were there specific user research methods that worked well?
		What specific design elements or features were implemented in the AI advisor to help reassure users that they were 'doing it right'?
		Were there any challenges in balancing transparency, trust, and the complexity of AI systems? How did your team address those?
		How did you measure whether users felt reassured or trusted the AI system?
		Were there any unexpected findings or user behaviours during the project that influenced your approach to designing trustworthy AI?
		Given your experience with conversational AI, what advice would you give for building trust in AI services in the public sector?

Interview Guide 1

Category	Clara / Anna	Research Question	Interview Question
Introduction	Both	What is my thesis about?	I'm working on my Service Systems Design master's thesis with the Municipality of Copenhagen's Citizen Service Development department. As I mentioned on LinkedIn, the municipality is developing AI solutions to integrate into public services, both internally and externally. One of these tools is now ready for gradual rollout to front-line employees. It is an AI-powered search-bot designed to replace the internal search engine, helping employees quickly find answers to complex citizen queries. A key challenge with this rollout is getting end-users to trust these AI solutions enough to use them. I am therefore investigation how we can design AI solutions that foster trust in the end-user.
Purpose	Both	What is the purpose of this conversation?	As you work closely with AI, I'm really looking forward to hearing about your experience with trust and AI. I'm especially interested in how you build trust and reassurance into the design and how that might compare to the public sector.
Duration	Both	How long is this conversation?	I have about 7-10 questions regarding your experience, so I expect this conversation to last about an hour.
Consent	Both	Consent for recording and handling of data	It would easiest for me to record this conversation to refer back to later. Would you be alright with that? If not, I can take handwritten notes.
Context	Both	Context to role and experience	Can you tell me a bit more about your role at SimCorp and give me some general information on any AI-solutions you may have worked on here?
Framing	Both	How is trust framed in a high-stakes industry	How do you define "trust in AI" in the context of financial services?
User related	Clara	Identification of trust drivers	Have you had any experiences of end users distrusting AI-solutions? What factors contribute to their trust or distrust?
Design	Clara	Exploration of the designing of AI systems	What is your take on "transparency" and "explainability" in AI? How do you balance them with usability?
Design	Clara	Exploration of the designing of AI systems	What design strategies have you used that been effective in building trust in AI services?
User related	Clara	Iterative design processes to address trust issues	What role does user testing play in shaping AI-powered services? Have you seen trust levels change during testing?
Communication	Clara	Setting realistic expectations	How do you communicate AI's limitations to users without reducing their confidence in it?
UX/UI	Anna	UI/UX techniques that enhance creditability	How do you design the frontend of AI-powered tools to make them feel more trustworthy?
Research	Anna	Comparison of findings	What were the biggest trust barriers you identified in your research?
Research	Anna	Learning from experience	Looking back, is there anything you would have done differently in your research? Any gaps you wish you had explored more?
Research	Anna	Referral to relevant sources	Are there any key papers or resources you found particularly valuable when researching trust in AI?

Interview Guide 2

AALBORG UNIVERSITY

DECLARATION OF CONSENT

Aalborg University (AAU) is the data controller in connection with my project.

This project is about exploring the design of trustworthy AI in collaboration with the department of Citizen Services Development at the Municipality of Copenhagen. In relation to this project, I want to understand how AI solutions can be designed to foster trust in the end user, in addition to the key insights and challenges from your experience. I, therefore, request your consent for capturing photographs and a voice recording during this time. These will be used for documentation purposes only, and will not be redistributed. I might use quotes and photographs from the interview in my project report.

You are free to withdraw your consent at any point throughout the interview or thereafter, with no obligation to explain the reason.

By signing this declaration, you consent to let me:

- Capture photo documentation and voice recordings of you during the interview

You always have the right to change your consent. If you wish to change your consent later on, you can revoke the consent by sending an email to: jsampa23@student.aau.dk

I will take steps to ensure that these images and the recording are used solely for the purposes they are intended, and are stored securely. If you become aware that the documentation is being used inappropriately, you should inform me immediately.

The General Data Protection Regulation entitles you to obtain information that you find under this link: https://gdpr-info.eu/

☐ I hereby consent to the use of photos by AAU in accordance with the above purposes and information. [please check the box]

Date:

Name:

_____

Signature

SAMTYKKEERKLÆRING

Aalborg Universitet (AAU) er dataansvarlig i forbindelse med mit specialeprojekt.

Projektet undersøger, hvordan AI-baserede services kan designes, så de skaber tillid hos brugerne. Det sker i samarbejde med Borgercenter Udvikling (BSU) i Københavns Kommune.

I forbindelse med mit projekt afholder jeg en workshop med formålet at teste og videreudvikle et værktøj (Trust Design Canvas) og en tilhørende playbook.

Under workshoppen vil jeg:

- Tage noter,
- Optage lyd (kun til transskription og analyse), og
- Tage enkelte billeder til dokumentation og evaluering.

Materialet vil udelukkende blive brugt til dokumentation og vil ikke blive delt offentligt. Udvalgte citater og billeder kan indgå i mit speciale, som afleveres på AAU.

Deltagelse er frivillig. Du kan til enhver tid trække dit samtykke tilbage – både under og efter workshoppen – uden at skulle give en begrundelse. Hvis du trækker dit samtykke tilbage, sletter jeg alt relevant materiale.

Ved at underskrive denne erklæring giver du samtykke til:

- ☐ Jeg giver samtykke til, at der må tages billeder under workshoppen
- ☐ Jeg giver samtykke til, at der må laves lydoptagelse under workshoppen

Du har altid ret til at ændre dit samtykke. Hvis du ønsker at gøre det, kan du kontakte mig på: jsampa23@student.aau.dk

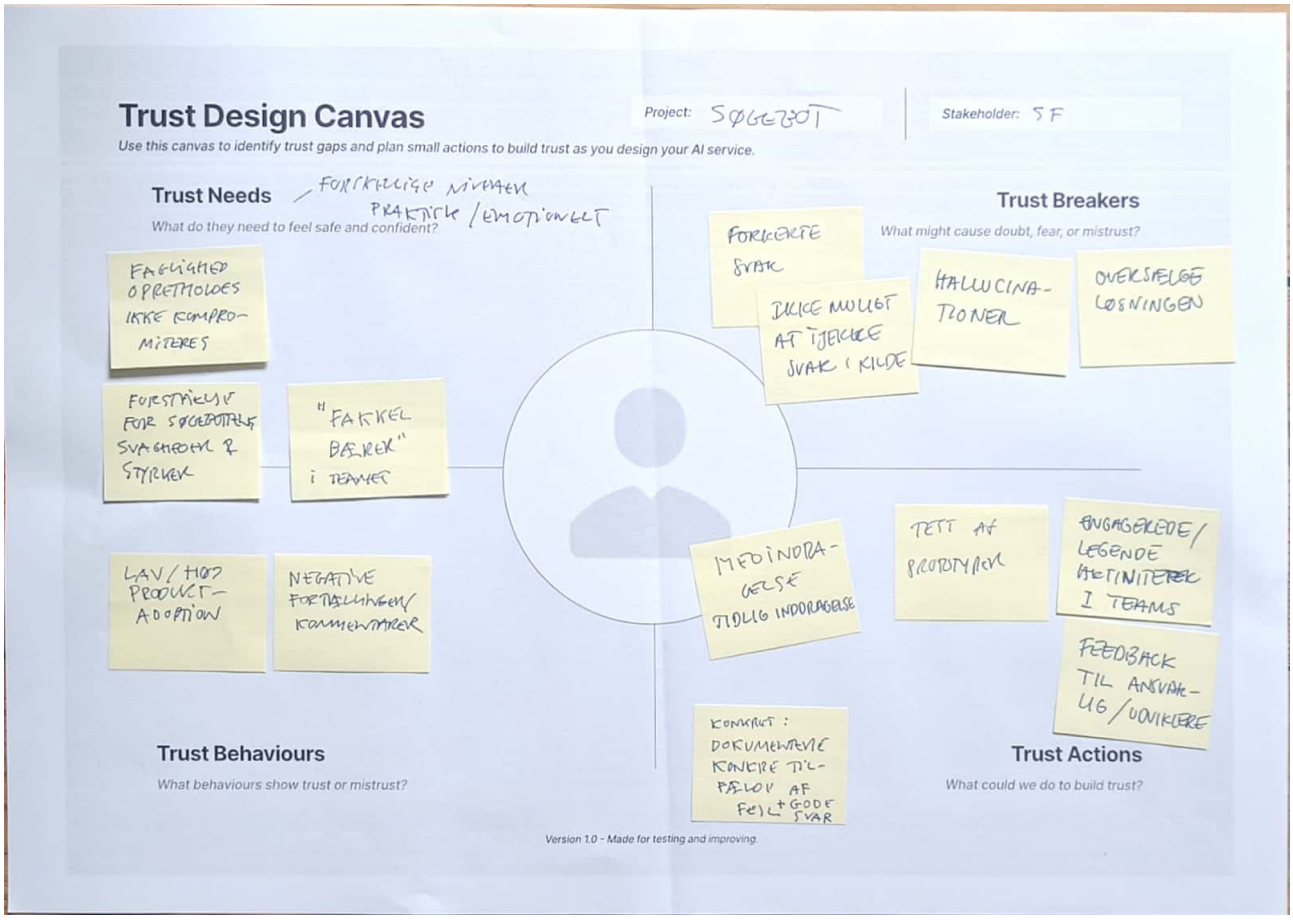
Jeg vil sikre, at alt materiale bliver opbevaret forsvarligt og kun brugt til det beskrevne formål. Hvis du opdager misbrug, bedes du kontakte mig straks.

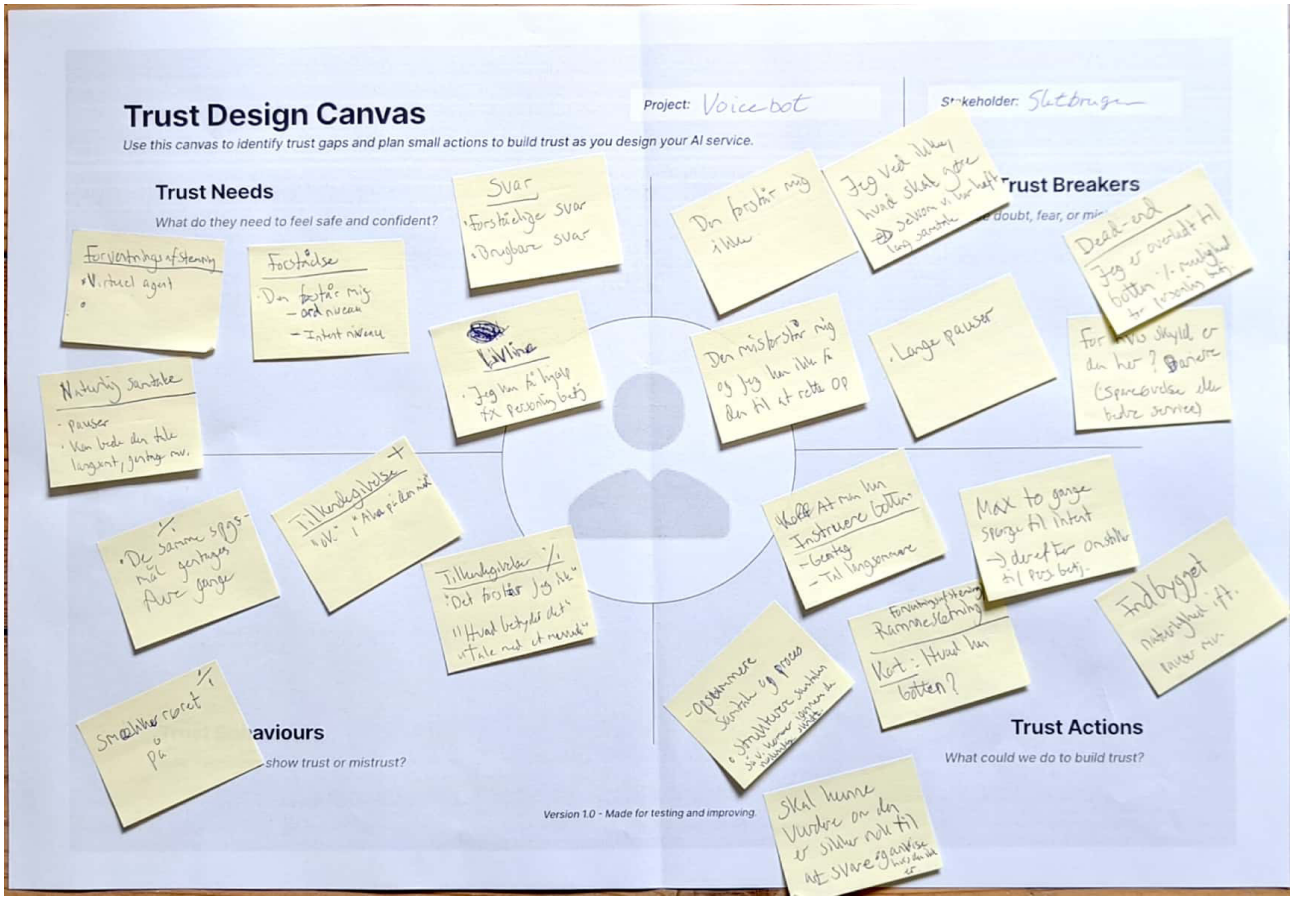
Læs mere om dine rettigheder under GDPR her: https://gdpr-info.eu/

Dato:

Navn:

Underskrift: _____





Trust Design Canvases completed by workshop participants