Mutuality in AI-enabled new public service solutions

E. Koskimies[a]* and T. Kinder[b]

[a]*Faculty of Management and Business, Tampere University, Tampere, Finland;*

[b]*Faculty of Management and Business, Tampere University, Tampere, Finland*

* Emmi Koskimies, Doctoral Researcher, Tampere University, Administrative Sciences, Faculty of Management and Business, Kanslerinrinne 1, Pinni A, 33014 Tampere FI, Phone: +358 50 318 2623, emmi.koskimies@tuni.fi. ORCID: https://orcid.org/0000-0001-6762-2371,

Tony Kinder, Visiting Professor, Tampere University, Administrative Sciences, Faculty of Management and Business, Kanslerinrinne 1, Pinni A, 33014 Tampere (+358 (0) 294 52 11 anthony.kinder@tuni.fi. ORCID: https://orcid.org/0000-0003-2769-3655

Mutuality in AI-enabled new public service solutions

This article explores the policy background of AI in ethical public services by applying the perspective of mutuality and trust to give a better understanding of the ethical evaluation of AI-enabled public services. The findings from Finnish government services emphasise that establishing governance rules arising from mutuality is not often viewed as a precursor to ethical evaluation, which was done post-facto with limited user engagement. We conclude that structured by a social mutuality framework, mutuality requires a systemic approach to ethics and active user engagement, which in turn requires an investment of time and cognitive attention by all agents involved.

Keywords: Artificial intelligence; mutuality; trust; public services; framework

Words: 9000

1 Introduction

While artificial intelligence (AI) offers opportunities to exploit digitalisation, claims of a transformational shift (Pencheva *et al.* 2018) or fundamentally changing economy and society (Ulnicane *et al.* 2020) sound hollow if AI imposes big-tech values on public services or society at large, as Hickok (2020) and Bryson and Theodorou (2019) argue is possible. As Kaufmann (2021) says, tech companies "making money at the expense of societal values, people, and the planet" breeds distrust in public organisations and civil society. Public services (meaning services designed and delivered by formally constituted, tax-funded public agencies, such as central government organisations) increasingly involve the private and third sectors. Thus, the boundary of power and control of what constitutes public services is porous, meaning market and non-market principles influence ethical principles governing public services. As Bryson (2019) notes, recognising this shifting boundary has important implications for who controls the controllers of AI. One-way isomorphic change (Weber 1978), with public sector imitating private sector, is increasingly unacceptable. An approach that diverges from the new public management (NPM) market predomination and closed, business-led epistemic communities (Haas 1992) is needed.

To maintain trust, public organizations need to be aware of the coercing forces in the environment and simultaneously solve organizational and ethical challenges in addition to technological issues (Mikalef *et al.* 2021) involved in the service design process. Black box technologies, meaning understood inputs and outputs with not understood transformation processes (Rosenberg 1982), such as AI, feature a steep learning curve and require conscious interaction between developer-experts and users. Current AI adaptation frameworks in public sector recognise the threat of technology-push and unethical practices (see Floridi *et al.* 2018; Bietti 2020; Ulnicane 2020; Nemitz 2018). However, prior research debating responsible AI adaptation has focused on technical aspects (Mikalef *et al.* 2021; Floridi and Taddeo 2016; Ananny 2016; Lin *et al.* 2012), vague ethical principles (Jobin *et al.* 2019; Fjeld *et al.* 2020), regulation (Rahwan 2018), or policies (Bryson *et al.* 2017; 2018). Few insights have emerged regarding the public management in organizational deployment of AI (Mikalef *et al.* 2020) and the actions of agents in multiagent collaboration to support trustworthy outcomes (Rossi 2018). De Sousa *et al*. (2019) also argue that AI requires new investigations about ethical implications.

Building on this gap we use the idea of mutuality to explore the policy background for AI in ethically acceptable public services. In doing so, we discuss current literature on mutuality, trust, and nature of AI in public service provisioning. Mutuality debate has taken place in psychology, sociology, and organisation studies (Dabos and Rousseau 2004; Jordan 1986; Blau 1964; Thibaut and Kelley 1959). Yeoman (2019) uses the term ethical organising debates, which we build upon. However, we do not see mutuality primarily as a shared psychological relation or meaningfulness in social exchange. Instead, we relate mutuality to boundaries of power and control problematised by crossing governances. We argue that mutuality comes from practical services delivered in ways acceptable to both users and providers, and it presupposes power arrangements taking user views of ethical acceptability and desirability into account.

Our theoretical contribution to public management research is two-fold. Firstly, we interpret mutuality in terms of power and control and expand current psychological and behavioural approaches to the public management discussion. Mutuality offers a way of exploring how AI-enabled innovations in public services can build trust by explicitly recognising that the boundary of market/societal control represents different values systems that each need respecting. Secondly, we suggest a new conceptual framework using the lens of mutuality and trust for approaching an ethical evaluation of AI in public services. Our objectives are to analyse data from developing AI initiatives in Finland adding to understanding how mutuality and trust influence ethical evaluation of AI-enabled public services. Our empirical contribution is a dataset of Finnish central government civil servants, private developers, and funders. We focused on public agencies involved in AI development in government services as they represent important public service agencies, and national government has adopted the goal of becoming a center for ethical use of AI (MEAE 2017). Following constructed grounded theory (Charmaz 2006), our framework elucidates how decision-making is influenced by mutuality and trust. Our research question, designed to take a more in-depth approach to the AI development process, is:

What actions are AI developers undertaking to achieve mutuality in the deployment of AI in new public service solutions?

Paper proceeds as follows. Our literature review analyses relevant research on organisational deployment of AI and the activities of stakeholders involved in service design. We introduce

mutuality and trust as a suitable lens to explore the actions of AI developers. After building a conceptual framework on these discussions, we present our methods followed by empirical analysis. After the findings, we propose theoretical conclusions and carefully outline generalizable lessons for public agencies implementing AI-enabled service solutions.

2 Literature review

2.1 Ethical artificial intelligence in public services

Public services are facing the challenge of steering fast development of technology and the transition from current ICT to AI (Floridi *et al*. 2018; Bryson 2019). AI is used as an overall term to describe a mix of computational methods and technologies, systems, that function in an intelligent fashion (Stone *et al*. 2016; AI HLEG 2019). Processes include learning, reasoning, and self-correction (Gillath *et al.* 2021). AI here is characterised by natural language processing, robotics, and machine learning using proprietary algorithms, referencing a range of public and private databases, analysing patterns in digitised big-data, resulting in services (Kinder *et al.* 2021). AI is already widely applied in public services (Dignum 2019; Coeckelbergh 2020) for example in data governance (Gupta 2019), decision-making support (Kankanhalli *et al.* 2019; Spieth *et al.* 2014), automation of practices (Kuziemski and Misuraca 2020; Chun 2008; Collier *et al.* 2017) and improving interaction, citizens' experience, and user involvement in service design (Kreps and Neuhauser 2013; Androutsopoulou *et al.* 2019). Overall, Wirtz and Müller (2019) argue, that the ambitious potential of AI is to improve the magnitude, speed, and accuracy of information and case processing, efficiency and effectivity of human labour, and downsizing of bureaucracy that can lead to improved public services.

AI has many potential benefits, but recent studies and practical cases show that uncontrolled transition can lead to misuse of AI and unintended consequences (Greene *et al.* 2019; Brundage *et al.* 2018). Injustices in the use of AI, inadequate legislation (tail-ending practice), failure to address ethical dilemmas by AI promoters, and governance clashes are widely criticised by Cath *et al.* (2017), Edwards and Vaele (2018), Roberts *et al.* (2016) and Ulnicane *et al.* (2020). In AI initiatives socio-technical ethical concerns arise from misguided, inconclusive, and inscrutable evidence, unfair outcomes, transformative effect, and traceability (Trocin *et al.* 2021). Failure to address concerns and dilemmas affects performance, security, control,

economic, societal, and ethical risk (Gupta *et al.* 2021). Often AI misuse cases combine these ethical concerns, for example, O'Neil (2016) and Eubanks (2017) detail practical injustices to different groups of people in large-scale predictive analytics deployed in public services using biased algorithms and/or datasets.

As Trocin *et al.* (2021) argue, there is a need for responsible AI research based on the limited understanding of important issues that emerge with intelligent technologies. In practical terms, AI, like other technological innovations can be evaluated using a combination of the ethical approaches on offer. These can be normative theories such as deontological ethics, virtue ethics, pragmatic ethics or consequentialism, or different branches of techno-ethics offering approaches to specific technological applications, for example, information ethics (Floridi 1999), data-ethics (Floridi and Taddeo 2016), computer ethics (Moor 1985; Johnson 1985), biotech or nano-ethics (Schummer and Baird 2006; Hunt and Mehta 2013). Mittelstandt *et al.* (2019) argue that solving ethical issues and mitigating distrust risk in AI requires raising awareness using general ethical principles around trust, privacy, and transparency. Numerous ethical guidelines with high-level principles address ethical issues of AI use (Hickok 2020) but critics view vague principles as idealism since they fail to explain how developers organize around AI initiatives (Mikalef *et al*. 2021). Also, such principles fail to guide the action of software developers (McNamara *et al.* 2018) or provide clear recommendations on normative and political issues (Hagendorff 2020; Mittelstandt *et al.* 2019).

Already numerous practices and procedures (socio-technical systems, professional practices, institutions) support ethical evaluation of AI (AIHLEG 2019), yet to address a rising public distrust of AI our research discusses the actions and morality of public agents and stakeholders involved in AI-enabled service system design. As Schaefer *et al.* (2021) argue, context influences the adoption of AI in the public sector, which also affects ethical evaluation. This means that the primary task for AI adoption differs between cases and influences the definition of goals, activities, agency roles, procedural guidelines, and codes of conduct. Ethical evaluation becomes situated and relational (Tännsjö 2002) which means that we can learn from different techno-ethical models but not apply them without re-contextualisation. The point of relational ethics is to avoid evaluations giving primacy to the technology. In Tännsjö's (2002) relational ethics approach interaction between facts, principles, and decisions varies with context and culture, as do ethical evaluations of technologically enabled service solutions. No

universal checklist is available; instead, a customised mix is needed, suiting each context and culture, problem, and solution.

Instead of focusing on outcomes and ethical principles, our research steers the focus to policy background by asking about actions developers are taking to support mutuality to provide trustworthy AI-enabled public services.

2.2 Trust

Since public services take responsibility for vulnerable people, trust is essential. Yet, as Winfield and Jirotka (2018) note, AI systems raise important questions around trust, including public confidence in the use of AI decision-making systems. They call for actions by individuals in application domains and at the institutional level. EU addresses the issues of trust in *Ethics Guidelines for Trustworthy AI* (2018), emphasising the ability of AI systems to engender trust. Trust in development, deployment, and use are foundational in EU's AI ethics.

Trust is problematic since it is relational, cognitive, and affective, being built on a leap-of-faith (Nooteboom 2002): it is a willing acceptance of vulnerability to actions of others (Dietz and Den Hartog 2006; Möllering 2006) and as Hardin (2002) notes presumes mutual empathy. Trust cannot be disentangled from control since trust accepts the exercise of authority and even coercive control (Lukes 1974; Weibel *et al.* 2016). Six (2005) argues that trust always meets trouble: it is always tested. Only those in deep trusting relationships empathetically understand the actions of others when tested. Trust lost is difficult to regain. Six cites the Dutch proverb: *trust walks in slowly but rides away on horseback*. High-trust relationships can dispense with the protections offered by formal contracts, as Weibel and Six (2013) note; doing so heightens the consequences of betrayal.

Rossi (2018) argues that AI has created a problem of trust. Public administration needs to be able to build a system of trust in technology and in those who produce it meaning bias detection and mitigation abilities and explainability for AI decision making. However, this is challenging as the absence of trust is part of Stinchcombe's (1990) *liability of newness*. This is especially so where technological capabilities exceed public understanding (Rosenberg 1982), or Schattschneider's (1975) *mobilisation of bias* creates misunderstanding. Since AI systems have direct and indirect effects on citizens, Dignum (2019) argues that citizens need

to understand what AI is, is not, and how to make AI trustworthy. Creating trust involves social and technical constructs that ensure responsibility in systems developed and use in dynamic contexts.

Dietz (2006) and Dietz and Den Hartog (2011) distinguish between trust as belief, or a trusting decision, and trust-informed actions. Logic-of-practice may serve to support or dismantle trust. Trust suspends uncertainty. Möllering (2006:111) argues a deepening of trust in AI relies on initial system performance and its explainability to all citizens or the representatives they trust. Ryan and Deci's (2000:99) self-determination theory helps: if AI respects autonomy, exudes competence, and deepens relatedness, it is deserving of trust. Alternative scenarios unraveling trust and creating distrust are apparent. De Bruijn (2007) argues, result in imposed performance control systems bereft of trust. For our purpose trust is a process of trust-as-a-relationship in which active citizens engage in *all* AI-enabled service design decisions.

Just as in any unequal relationship (such as parent-child) the onus is on the person with power and knowledge to explain and answer 'why' questions, using mutual relationships to build trust. From the citizen's viewpoint, effort in active learning and engagement with design processes are necessary conditions for building trust in AI. Trust requires active relationship building and learning by all parties: mutuality.

2.3 Mutuality

Research on AI has established that robust implementation mechanisms to support trustworthy AI can only be done using holistic, multi-disciplinary, and multi-stakeholder approaches, ensuring issues are identified and resolved in a cooperative environment (Hickok 2020; Rossi 2019; Watson 2014; Barret and Baum 2017). Technologically enabled innovation is future-oriented and therefore often constrained by heritage context, cultures, and ways-of-working (Wartofsky 1979; Bernstein 2000; Daniels 2016). In a multi-agent environment, agents bring together different decision systems, institutional logic, and information flows that can easily conflict (Rossi and Tuurnas 2021). Especially, if inter-working between diverse governance arrangements (e.g., market vs free public services) is unresolved. To address the need for an interdisciplinary and cooperative approach we use mutuality. Mutuality is a form of governance, in this case suggesting agent interdependency featuring trustful relationships as opposed to (for example) purely market governances in which for-profit principles mediate

decisions. Governance here is deployed in a wide sense of rules and norms guiding decisions and actions (Kinder *et al.* 2020) and includes mutuality between agents in public service development.

Most theorisations of mutuality are psychological or clinical (Jordan 1986) debating individual envisioning of dyadic, triadic relationships such as medical confidentiality (Henson 1997), parent-child relationships (Tronick *et al.* 1977), or love commitment (Drigotas *et al.* 1999). We find these approaches limited as our analysis focuses on multi-agent cooperation where agents intend to create new service solutions by negotiating new team governance, language, and ways-of-working, setting aside dyadic relationalities. In the context of service design, Yeoman (2019) argues that mutuality is a *development organising principle* guided by meaningfulness and wellbeing. Understanding values, principles, and practices as part of meaningfulness can be helpful in conceptualizing relationships in AI development. Especially trust, respect, honour (Nietzsche 1997), and emotional attachment (Vygotsky 1934) between agents as projects bring together stakeholders from diverse disciplines and governances (Watson 2014; Barret and Baum 2017).

Organisational studies research also addresses issues of mutuality. Thibaut and Kelley's (1959) interdependency theory recognises mutuality of dependence: are partners dependent equally on each other's behaviour to achieve the desired outcome? Similarly, Dabos and Rousseau (2004) tie mutuality to psychological contracts where mutuality is an agreement on what each party owes others. We find Blau's (1964) social exchange theory more helpful in forming mutuality as it seeks to understand trust and acknowledges boundaries of power and social dominance. For AI ethics this is important as it seeks to answer the questions *by whom, how, where, and when will positive or negative impact be felt* (Floridi *et al.* 2018). What is technically possible may not be desirable or useful: how do we evaluate usefulness and ethical acceptability? Following Weick (1995) our approach is that exploring organising is more revealing than studying organisations. Especially so for AI development because AI initiatives often necessarily disturb existing hierarchies and power distributions.

Our framing of mutuality is socio-economic and related to power and control since social control issues already exist in conventional technologies (Bryson and Kime 2019). Especially when providers are partnering with major tech companies (Friedman and Nissenbaum 1996; Coeckelbergh 2020). Researchers point to implications for information asymmetries between

consumers and policymakers, weakening accountability, democracy, and regulatory standards when using technologies shaped and provided by global oligopolies (Nemitz 2018; Greene *et al.* 2019; Gasser and Almeida 2018). Ethics turn into *warm words* and abstractions in Bietti's (2020) concept of *ethics washing*: the instrumentalisation of ethical language – use of ethics "*as an acceptable façade that justifies deregulation, self-regulation, or market-driven governance*". Market principles guide AI development towards the lowest cost and highest profit margin, whereas mutuality-based development is driven by agent satisfaction with service effectiveness, especially for users. An example of non-market mutuality is the Finnish Linux network (Castells and Himanen 2002), later Linux product (Red Hat). Mutuality includes giving without expecting reciprocation and contributions for the greater social good. Thompson's (1971) *moral economy of the commons* (see Hardin 2002) is an example. Mutuality is unmediated by price and commodification.

Non-market (sociable) mutuality requires active relationship-building found in the teacher-learner relationship, volunteering, or mutual-aid society. Osborne *et al.* (2015) argue it places users at the center of public-service system relationships. Market mutuality hides power in customer-provider relationships, what Nietzsche (1997) terms *formal equality*, often lacking respect, honour, and trust. Social mutuality shares and respects diverse opinions, whereas market mutuality consults and configures around marketing preferences while retaining control. Therefore, mutuality-based development differs from market-driven processes in terms of knowledge flows, levels of trust, and time spent on understandability.

2.4 Towards a social mutuality framework

Having rejected supposedly universally applicable normative evaluation of AI (see Lin *et al.* 2012; Ananny 2016; and Floridi and Taddeo 2016), the framework we use to structure our findings and analysis focuses on the processes of trust-building. Framework acknowledges AI's complexity and activates agents in the relational processes designing and creating AI-enabled solutions to citizen's problems. This is a much deeper engagement than post-facto judgement using abstract ethical principles as in most AI ethics debates (see Jobin *et al.* 2019; Fjeld *et al.* 2020), without any clear result (Hickok 2020). Our mutuality framework contributes to public management theory by synthesising the conceptual approaches to trust and mutuality discussed above drawing attention to four factors: a) context and culture, b)

problem framing, c) knowledge flows and d) organising, the interaction between these supports ethical AI design.

- *Context* refers to professional AI implementing policies, structures, procedures, standards, rules, budgets: 'hard' features of the development environment (e.g., Schaefer *et al.* 2021). 'Soft' features of the design environment i.e. *culture* are predispositions, ways of working, and social meanings in organisational, occupational cultures (Wartofsky 1979; Bernstein 2000; Daniels 2016) and the wider Finnish culture. Agents in the design environment (often AI-technicians, citizens, and formal public service providers) bring multidisciplinary knowledge and emotional attachments to find shared meanings and goals, all informed and shaped by the context in which a new service solution is emerging Rossi 2018; Watson 2014; Barret and Baum 2017). Thus, in each AI-enabled public service context, an acceptable mutuality and relational ethics are negotiated (Tännsjö 2002).
- *Problem framing* features a long-term vision, framing and scope agreed upon between agents involved in the development process (Floridi *et al.* 2018). Agents need to be clear about the customer profile, problem, how new AI-enabled solution will perform differently, and the capacities and purpose of AI. Re-framing is likely to occur as contextual limitations become clarified and as technical and ethical constraints emerge (Trocin *et al.* 2021).
- *Knowledge flows* include identifying relevant user groups and agents in the public, private, and third sectors participating in a service design that solves citizen's problems within prevailing institutional logic and achievable inter-organisational delivery parameters. Knowledge flows include 1) providers and citizens understanding the limitations and possibilities of AI (Dignum 2019) and 2) AI technical agents understanding the emotional commitments of service users and providers. A two-way knowledge flow based upon mutual respect 'pulled' by a shared commitment to creating a better and workable new service solution.
- *Organising* features all actions: operational practices, management, governance, and decision-making structures (Weick 1995). Actions taken to achieve AI strategy in place (international, national, local, organizational) (see AIHLEG 2019) and identifying a situated ethical toolkit suitable for the particular problem facing the project, including evaluation standards.

Our framework, figure-1, envisages interaction between these four factors, with problem-focus, learning, and mutuality at the center. Each potential solution will be tested against these three standards: (a) does the proposed solution solve citizen's problem, (b) is mutual learning embedded in the proposed solution, and (c) is mutuality acceptable? Learning is the oil in the machine – creating trust and respect for the technical, effectiveness, and efficiency standards and affective commitments of other agents. Framework envisages the design group iteratively going around and around these factors until a consensual, perhaps compromise, new services solution, and associated ethical standards are agreed upon. It is a dialogical process characterised by shared language, concepts and understanding of context and culture.

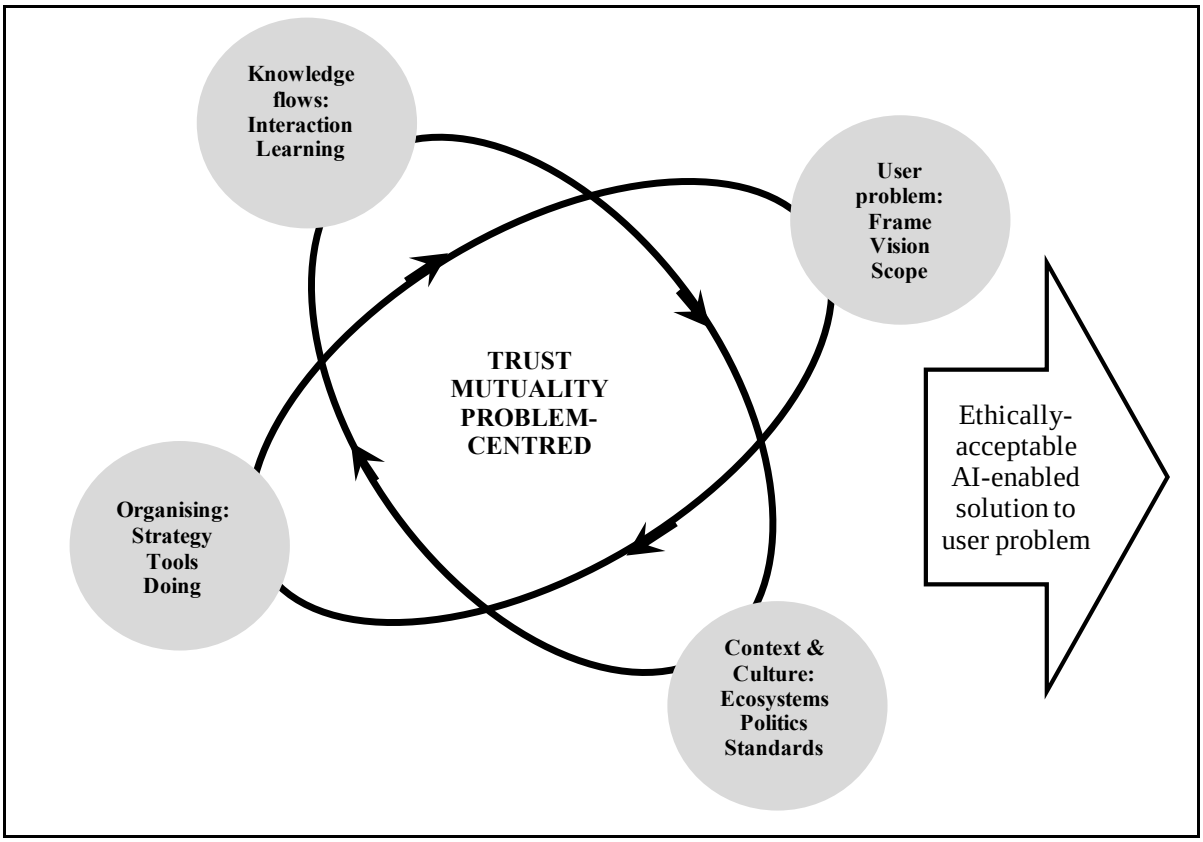


Figure 1: Framework: mutuality and trust influencing the policy background for AI innovation in public services.

Conceptual framework invites policymakers, managers, technical and business partners with front-line staff and citizens to negotiate the boundaries of mutuality and trust in creating a new AI-enabled service solution. The solution embeds every stage, design decision, a consensual view of ethical compliance, including possible compromises and trade-offs (an example could be between efficiency and access or accuracy). Framework encourages mutuality, understandability, and explainability within new service solution design processes, referencing

diverse ethical approaches while citing standards and regulations. It guides stakeholders to evaluate control and power shifts, for example from technical providers (does it work?) to citizens (does it ethically solve our problem?). As Bietti (2020) the framework encourages discourse around social wellbeing issues – why and how new service solutions become recommended. It encourages sharing a multiplicity of cognitive-emotional moral evaluations that pull individuals towards solution options and invite different individuals with varying stances to integrate their evaluations (Kaplan 2014), crossing cultures, including professional, gender, and ethnic cultures.

3 Method

Context of empirical study

Finland is a Nordic welfare state and high-trust society (Fina *et al.* 2021) that MEAE (2017) calls to become a European center for ethical AI. The country can be described as a pioneer for digital developments. Finland a heritage of world-leading software clusters, such as Nokia, and now advanced software sectors that support the capacity in the public sector.

Public sector AI development features mainly pilot projects in central and local government (for example in transport, energy, security, health and social care, and payment systems). Projects are supported by national initiatives such as Aurora-AI program which encourages AI innovation and new service chains based on transitional life events (e.g., family circumstances). Programmes support nationwide cooperation and building multi-stakeholder ecosystems that flexibly interact (SAIP 2019) and enable mutual sharing of information, compatibility protocols, and platforms for cross-governance development. Finland's culture supports public-private collaboration and problem-centricity which addresses AI-expertise deficit in the public sector while providing data and expertise from service models lacking in the private sector.

Finnish central government service development, the context of our empirical study, is still at the early stages of adopting AI technologies to overcome a growing demand for basic services caused by for example Covid-19 and demographic changes.

Research design

Our research is exploratory; variables, boundaries, and causalities for AI use have yet to be agreed upon, defined, or evidenced by the contestation over definitions (of AI), mutuality, and

trust and how ethical principles apply to using AI in public services. It is a realist (Fisher 1988): we construct a plausible interpretation of how technical and social aspects of AI blend in practice. Our data is qualitative; following Easterby-Smith *et al.* (1991), we delve for meanings from data relating to social interactions. Our unit of analysis is the policy background facing AI innovators in public services. This policy-level unit of analysis dictates the data needed (policy-level and partners) and thus, this unit of analysis excludes service users, whom we would expect to meet in research on service design and the associated design of ethical evaluation.

Research question

Our research question relates to the gaps in the literature analysed above and the need to clarify mutuality and trust as they influence the ethics of AI. AI requires new investigations about ethical implications, especially in a holistic, multi-disciplinary, and multi-stakeholder environment (Rossi 2018; de Sousa *et al.* 2019). Following Silverman's (2001) advice, we ask a processual question: what actions are AI developers taking to achieve mutuality in the deployment of AI in new public service solutions? The advantage of a broad question is a holistic perspective, the disadvantage is the need to condense the evidence trail. Charmaz (2006) draws attention to misinterpretations and fuzzy meanings while allowing the emergence of new factors and causal relationships. We emphasise the critical importance of context and culture for learning and ascribing meanings, a further justification for a broad research question.

Data gathering

Our dataset comprises ten interviews (from spring 2020) under the auspices of ETAIROS project, each lasting one hour. The sample (Figure 2) comprises five senior policymakers (I-2, I-3, I-5, I-8, and I-10), two private-sector AI developers (I-4 and I-9), two private-public AI innovation funders (I-6 and I-7), and a university AI expert partner (I-3). We used purposive sampling (Etikan *et al.* 2016) to find sufficiently experienced people in AI innovation in the public sector to comment/narrate on AI development and thereby support answering our research question.

Key actors		
I-1	Male	Digital population and data service agency
I-2	Male	Ministry of Finance

I-3	Male	Finnish Center for Artificial Intelligence (FCAI)
I-4	Male	Technology and software provider
I-5	Male	Prime Minister's Office
I-6	Female	The Finnish Innovation Fund SITRA
I-7	Male	Finance Finland (interest organisation)
I-8	Male	Ministry of Finance
I-9	Male	Aalto University / Nokia Bell Labs
I-10	Female	Ministry of Finance

Figure 2: Interviewees: code, gender, and organisation.

We used an open-ended questionnaire divided into five themes: (1) introduction and use of AI in public services, (2) Goals for use and effectiveness of AI, (3) Guidance and regulation for AI, (4) Collaboration with stakeholders, and (5) Values and ethics. Interviews were conducted by Author-1 and ETAIROS colleagues. Interviewees gave written consent and anonymity was guaranteed. Following constructed grounded theory conventions (Charmaz 2008), the questionnaire left open choices of terminology and sequencing to the interviewees. We, therefore, avoided the explicit use of terms such as trust and mutuality in our interviews, allowing the interviewees to choose the terminology. This could make internal validity problematic. However, we took care in theming and triangulation to seek interviewee meanings and not rely on word matching.

Research strategy and analysis

For exploratory research, we followed Charmaz's (2000; 2006; 2008) constructed grounded theory (CGT) approach. Themes from previous research were embedded in the Figure 1 framework to guide data gathering, presentation, and analysis, and referencing new data and meanings. Analysis of the data proceeded according to Charmaz (2006) by first identifying patterns and initial codes, second reducing them to focused codes, and third, in the advanced theoretical coding, we built storylines. An example of the process is provided in Figure 3.

Basic statement	**Initial code**	**Focused code**	**Theoretical coding**
"But one conflict comes when technographics develop AI. So, professionals and regulators and officials are the ones that think for the people." (I-6)	Lack of user engagement The value of citizen involvement Dialogue Technology push	Knowledge flows	Social mutuality framework

Figure 3: Example of the coding process in CGT

Self-reporting bias is a problem in all interviewing. Following Yin (2014) and Charmaz (2008), we took care to validate responses concerning previous research with which we triangulated, including our previous research on AI and public services in Finland (cite authors), seeking internal validity. We discounted quantitative mixed methods since this is an emerging area in which established definitions and variables have yet to form. We considered using focus groups. However, due to the nature of our sample, individual interviews offered greater data-gathering opportunities.

Following Bryman's (2004) and Yin's (2003) advice on the conduct of social research allowed us to check internal validity. The generalisation from our research would need careful re-contextualisation since context and culture influence processes and relationships in each ecosystem. The main limitation of the research is its rootedness in the Finnish context and culture (including institutional arrangements and knowledge distribution systems), meaning great care is needed before transferring the results.

4 Findings

Part of our purpose is to offer a new analytical framework for use in the emergent area of applying AI-enabled solutions to local public services. We, therefore, follow Kinder (2002) and Yin (2014) in presenting our data thematically, structured by the four factors derived above from our literature review and embedded in our analytical framework: problem, knowledge, organising, and context and culture.

4.1 Problem framing

The interviewees pointed to *an expectations gap* as to what is achievable: practical applications of AI are still simple (I-7). I-10 referred to fears, dystopian myths, and sci-fi fantasies about AI polarising perceptions and the unrealistic image of AI influencing R&D funding. I-8 emphasised AI's potential while agreeing that the *management of expectations* is important. I-3, I-8, and I-10 said a lack of understanding of AI can cause missed opportunities or a preference for low-risk projects when an alternative is needed to 'traditional black box thinking' (I-10).

The purpose and aim for AI use were brought up by the interviewees. I-3 and I-6 referred to 'AI's fitness for purpose' and the need to carefully choose project problems. I-2 stated that AI should not be used for deep social and political solutions because machines cannot and will not understand human beings. The interviewees commented that inadequate problem-framing results in failed innovation, I-7 referring to 'these projects in which they do not understand what they are aiming for'. Although the interviewees mentioned customer service as an aim, few details were given: I-4, I-1, I-6, and I-10 mentioned efficiency improvement and I-7 and I-4 competitive advantage.

From a design point of view, *citizens and service users appear distant* and informant I-10 described citizen even as fictive, a filler term, someone as an end-user. Interviewees gave the impression of an undifferentiated, unsegmented set of users, and although human-centred AI was mentioned, it was not clear what it means in practice. One informant described how often technology came first but understood the importance of putting people first, which EU's AI guideline documents emphasise, but the problem is that 'we have not determined what human-centric means' (I-1).

No clear picture emerged of what the interviewees expected from service users, for example in terms of data literacy and how trust is built practically. I-8 and I-1 said actively involving users in the design was difficult; users do not understand final system design and choices. I-10 also questioned 'whom systems are developed for'.

4.2 Knowledge flows

The interviewees feared that a lack of communication and mutuality would create conflict, especially in complex services. They bemoaned *the insufficient dialogue* between stakeholders, developers, and end-users for ethical evaluation. I-6 felt that in the technology hype, it seems that everything is possible with technology, but agents do not include end users or professionals who know the substance in the process in time, and more dialogue is needed.

This is linked to how AI is understood in design and development projects by developers but also users. Interviewees worried that AI was accorded more agency than a tool. I-3 described how things often start to go wrong with trust-building when human agents are made invisible in the discussions. Focus on development shifts to technology, for example, algorithmic fairness, instead of thinking about organisations, agents who use AI tools, and the fairness of

the use, not only the fairness of one component. Public misunderstanding of algorithms was also frequently mentioned. For example, I-10 mentioned pandemic models, which have been the topic of conversation in the last few months: 'Only a few think that those [models] are based on machine learning that uses data analysis.

The developers argued that the government, public officials, and technology providers are responsible for making AI easier to understand. I-6 said the government should generalise its understanding of AI. However, some AI professionals hide behind the technical language. As mentioned, knowledge flows were also a question of *language* to interviewees.

The ethics discussion of AI they say is 'early stage' or 'still developing'. They pondered at what stage the ethics discussion was appropriate, most commenting that since projects were early-stage, ethical evaluation was 'not yet necessary', and only a few discussed ethics at the design stage.

> This is the chicken-or-egg problem… If we wait so long that the system is already coming or here then this discussion is too late. It then leads to the solutions being based on technology push or some commercial interest. (I-5)

Whereas the interviewees expressed exasperation with user inputs, the language suggested that AI was idealised as faultless, despite developers knowing that most programming comes accompanied by faults and bugs. 'Data biases and incompatibilities' were mentioned (I-9) as was using systems for purposes other than their original design, contributing to low mutuality and trust in outcomes. I-4 felt some performance problems were the result of 'bad communications', not technological failure. I-6 said that public discourse showed little understanding of AI and some developers were poor at explaining data choice and use. Standards, such as the Fair Trademark, help citizens gain trust.

The interviewees identified the benefits of multidisciplinary collaborations and broad ecosystems for trust-building in AI-enabled solutions, mentioning users, technology producers, and government organisations. User involvement was deemed more pronounced in NGO-led projects and especially those targeting marginal groups. Overall, the interviewees had found user collaboration important but difficult, and I-6 from SITRA described attempts to find a 'good functioning model for citizen collaboration' as often unsuccessful.

One reason for the lack of citizen collaboration might be that the interviewees differed on the value of citizen involvement, I-3 from FCAI says that citizens are poor at evaluating and expressing their needs, especially in projects creating new service systems of which they have no experience. The interviewees noted that a lack of user engagement leads to conflict and a potential loss of trust.

> One conflict comes when technographics develop AI. So, professionals and regulators, and officials are the ones that think for the people. (I-6)

Also, a risk of elite controlling AI development exists.

> We use such terms as empowerment and things that are usually available for people that already handle their business well and take advantage of the services available. But how about those who don't know that we offer something for them, how do we help those people? (I-2)

I-4 suggested a special effort be made to involve users in systems affecting minority groups, seeking to avoid what he called deterministic logic.

AI adaptability to multidisciplinary and complex service ecosystems works successfully, the interviewees said, provided development-minded service professionals support the innovation project. External pressure to find new ways of meeting user needs can be helpful, especially, I-7 said, if 'user trust in digital services is decreasing'. Lack of coordination between disciplines means that key design decisions are taken inside the government, especially where 'dialogue between the government and technology providers is weak' (I-3). According to I-4, the government too can fail to see the potential of AI or adequately resource projects.

Mutuality, the interviewees said, depends on finding a common language and shared understanding of the project environment and culture into which an AI-enabled solution fits. Overall, the interviewees believe public sector acknowledges limited know-how (*mētis,* tacit learning) and knowledge (formal), which explains its hesitation over AI rollout.

4.3 Organising services

Wary of AI hype, the interviewees discussed *incremental and radical transformational models* for AI adoption in the public sector, some arguing that the public sector lacks the courage or knowledge for a radical model, others, such as I-2, arguing that smaller public organisations lack the capacity. Public service providers stressed their duty of care and risk aversion and the pace of change being influenced by retaining public trust. I-8 noted that policymakers are

cautious when using AI since there is strong trust in officials in Finnish society. If things are done too fast and uncontrollably, it could potentially harm trust.

The interviewees emphasised whole system solutions, not those 'where every shop optimises itself when we should be serving the global epidemiological crisis' (I-4), acknowledging that larger projects pose issues of interoperability (data and hardware) and new coding. I-4 said that the lack of bigger development leaps is related to the inability of organisations and people to change. As a result, technology is shaped to fit the organisation. An alternative viewpoint offered by I-4 was that instead of shaping technology to fit use cases in organisations, we should think more about what the problem is that the technology should solve, in which cases we need AI and in which cases we do not need technology at all.

Multidisciplinarity poses challenges, since cross-sectoral procedures and activities are only forming in AI. Public-private partnerships shift responsibilities from the public to the private sector. I-8 questioned where future responsibility would lie. Commenting on the inter-organizational consequences of intra-organizational change, I-1 said that questions of decision-making responsibility have become vague in recent years, and in the open network collaboration the ownership of the project is often pushed to other agents.

The interviewees felt that AI ethics toolkits are at the early development stage. The main ethical challenges mentioned were data use (how and why), responsibility, fairness and legitimacy, and the increase or decrease of trust as an outcome. The developers criticised abstract ethical principles which are hard to translate into practical guidelines and wanted to avoid them being 'decorative, so as soon as the firm's financial profit is at stake, then they are the first thing to go' (I-3).

They cited as practice other central government organisations such as the Social Insurance Institution of Finland and Finnish Immigration Services, noting the need for impact assessments and legal standards. I-7 called for objective supervising agencies to help with ethical assessments. Recognising the user's subjective expectations and experiences of AI as important, the interviewees questioned how these connect with levels of trust and the acceptance of AI-enabled services.

> As people, we are more comfortable with the other person making the decision. But when a machine arrives at a wrong solution, it is voiceless, faceless, cruel, and wrong. Ethically, there is the dimension of feeling. (I-5)

4.4 Context and culture

The interviewees distinguished a narrow technical approach to user problem-solving from a wider socio-technical user-in-environment approach (which they favoured) reflecting context, opportunities, and constraints. I-10 described that fundamental differences in the views between stakeholders exist and the technology industry has difficulties in understanding from the administration's point of view why they want something.

They believed that AI hype makes user involvement more difficult and criticised technology providers pushing AI technologies regardless of the appropriateness or ethical suitability.

> We need more patience, the courage to be dull. A lot of firms function from their business logic and from that point of view cause the hype. This is no news because it's their advantage – short-term advantage. But the long-term advantage is more questionable. (I-1)

The interviewees felt some 'narrow' AI applications are examples of a poor fit in what are complex service environments, and in a small country, the database choice is limited for complex services. I-3 described AI's ability 'to anticipate an individual life's destinies and life happenings surprisingly badly' because of complexity.

Lock-in to existing mental models can be a danger, the interviewees believed. I-7 pointed to ethical dangers. They argued that AI technology is neither 'good' nor 'evil'; its value depends upon the use to which it is put, and the problem is reducing human biases.

The political atmosphere was cited as a context affecting the central government, requiring designers to work around political and financial exigencies, sometimes against their own political or ethical preferences. I-5 said, 'the political atmosphere changes as well as values', sometimes posing challenges for ongoing project funding and long-term thinking. There is no consensus on democratic principles yet, nor where to allocate money, even though 'we are at the point that we invest a lot of money in next-generation technology' (I-1).

The interviewees believed legislation trails technological developments, allowing developers to push boundaries. Standards were considered preferable to regulation, especially as families of technologies converge.

5 Discussion

Following the figure-1 framework structure, we explore mutuality influencing trust in service processes, triangulating with previous research.

Problem-centred

Previous research emphasises what Figure 1 refers to as the problem framing, vision, and scope agreed between innovation stakeholders as an essential starting point for successfully innovating a new service solution. However, the findings show that in Finnish government services, discussions with service users or their representatives were limited in practice, even though agents understood the importance, as they emphasised human-centric approaches to AI innovation. Users were often not involved in project framing. Also, the idea of the citizen (customer profile) was narrow or undetermined, and users were referred to as 'fictional' and 'distant'. This is an admission of the problematic process and subsequent difficulties in embedding the users' viewpoint on processes and outcomes.

Our unit of analysis (background to policymaking for AI innovations in public services) precludes user involvement limiting interviewees to public-sector policymakers, private-sector developers (AI technical), and public-private funders. Public-service users include both final (citizen) users and street-level bureaucrats, complicated (Watson 2014; Barret and Baum 2017) by often integrating decision-making between professionals from different service disciplines. Without these user groups providing input to develop trust in the projects (Weibel and Six 2013) and outcomes, the projects are left considering abstract (deontological) ethical principles.

The interviewees applied Rosenberg's (1982) black-box to AI technology; many referred to explainability, misunderstanding AI's fitness for purpose, and the overall understandability or unrealistic expectations – the absence of mutuality. Developers suffer from dissonance, emphasising the need for trust without trying to improve explainability/understandability for users: it appears to them as someone else's problem.

One major advantage of AI is the ability to interrogate multiple databases and service histories to create nuanced sets of users; machine learning can be used to move beyond undifferentiated users towards careful segmentation and then multiple routes to service satisfaction (O'Neil 2016). This benefit was acknowledged (for example, in identifying and engaging minorities), but the failure to deeply understand users prevents development projects moving into this rich

area of service design. In technology-led innovation, the absence of users means that the potential of technological innovation is limited. There is no coupling between push or demand-pull and little opportunity to trade-off between the different goals of stakeholders if some of them are voiceless.

Knowledge flows

A project is a specially contrived learning environment aiming to improve outcomes by altering processes. Unlike command-and-control hierarchies, AI projects involve a range of stakeholders from a range of governances and specialisms (Hickok 2020; Rossi 2019; Watson 2014; Barret and Baum 2017), making it difficult for a central controller to unilaterally dictate activity. Instead as an ecosystem, leadership is done by trust-building the power of ideas in the epistemic community (Haas 1992). The interaction around knowledge flows, shown in Figure 1, therefore results in governance clashes, (Ulnicane *et al.* 2020), clashing values and principles (Fjeld *et al.* 2020) and different reference sets of legislation and professional standards (Edwards and Vaele 2018), often with different sets of ethical dilemmas foregrounded (Roberts *et al* 2020). Resolving clashes over language and values encourages stakeholders to learn from each other, aligning proposed service solutions to the context and culture by intense (forming-norming-storming-performing) interactions.

The potential of knowledge flows are yet to be exploited, for example around the central government experience of AI; learning and communications between stakeholders are limited, though there is some collaboration with NGOs. The language shared between stakeholders present often relates to efficiency, speed, costs, i.e. the low-hanging fruit, more advanced big-data analytics than highly segmented users, and (machine-learning) routes to outcome satisfaction. These knowledge flows take the projects away from ethical user satisfaction with new solutions to service problems and instead towards managerialist goals.

In the absence of users, existing project delivery and decision hierarchies are left intact, as are inter- and intra-organisational hierarchies (Eubanks 2017) as well as centralisation of power where few control the decision making. From this perspective, the potential of AI is not realised: even in cost-benefit terms, the projects do not realise their potential, still, less do they result in solutions that they can say have stakeholder ethical endorsement. The return on a limited R&D budget is thereby limited. In short, narrowly defining the complexity of the service ecosystems narrows the knowledge flows, leaving users (SLBs and citizens) to post-

facto (yes or no) judge ethicality instead of designing-in ethical approval at all stages of design, including the emotional touchpoints so troublesome to final users (Radnor *et al* 2014). We see technical solutions to what may be (for the users) a social problem.

To summarise, knowledge flows and depth of learning are severely limited in many development efforts by the project design decisions, lack of shared language, technology idealisation, and power struggles between agents.

Organising

Importantly, Figure 1 frames organising (strategy, tools, doing) as an activity, particularly learning and innovating (Kinder 2010) activity, and not relations between organisations, giving a processual focus to the analysis. Numerous interviewees commented on the early stage some of these AI projects are at; other interviewees (I-4 and I-1) questioned at what point in project processes ethics should be considered. For Tännsjö (2002) and other ethicists, the answer is that a project that does not involve ethics-in-context from the beginning is doomed to limited achievement. Interestingly, instead of referencing what the user may think about ethical validity, the interviewees looked horizontally to previous Finnish projects (see I-3; I-7), considering the failures or successes to transfer technologies and processes (best practices perhaps without re-contextualisation) from the ethical evaluation of other projects – a denial of the importance of unpacking each point in the project process.

Some of the projects featuring injustices cited by O'Neil (2016) and Eubanks (2017) appear to have suffered from exactly this lack of attention to organising processes. Apart from referencing other Finnish projects, the interviewees referenced the need for a holistic or big system approach. In one way this is understandable, one public service links to others. However, the point these interviewees are making seems to be different; their interest is in a widely scoped technical AI system that reaches into a wide array of databases, algorithms, and machine-learned new technical processes (see I-1; I-10). Finland has many technological innovators (lack of designer's technological knowledge was not seen as an issue). Unfortunately, this easily leads to interest in technical systems and their adaptation, and only afterwards in their application to solve social problems and acceptable ethicality. It may be that this underlines some of the comments from interviewees (I-10) that somehow, ethical evaluation can occur without consulting users during the design stages. This differs from the findings Laitinen *et al.* (2017) made in which successful co-design in Finland involves users

and each design tests its social acceptability against a wider democratic footprint, an argument supported by Dignum (2019).

In summary, cross-sectoral activities are still forming, and ethical toolkits are in the early development stage in AI design and delivery. As technological knowledge is highly valued and mutual organising still emerging. Competence in organising AI projects and new service solutions lies on the developer's side of the society/market divide: mutuality involves following, not challenging the technical developers. However, the lesson from technical innovation research in the public sector is that mutuality requires recognising that both sides of this divide have a contribution to make. This contribution is not a post-facto judgement on ethicality, but instead a contribution to each stage of organising the AI project.

Context and culture

Context in Figure 2 refers to 'hard' aspects of the situation (structures, rules, budgets, standards), while culture refers to 'softer' aspects: predispositions, ways-of-working, and social meanings. Most previous research on AI-enabled new service solutions refers to the environment in which the innovation is occurring. Bietti (2020) emphasises the short-term nature of any processes for which politicians want to claim success; Hickok (2020) points out that, unlike the private sector, public authorities cannot choose customers and must design all services to feature equity and justice. Clarity of design brief is therefore critical for the inclusion of the diverse needs of users. Osborne (2006) points out translating usability into the legally defensible rules governments must make.

Making sense of clashes between different professional and occupational cultures in a multidisciplinary context takes time and patience (Dignum 2019). Often this begins with language as the terminology differs between professions. Kinder (2010) makes the point that the context may include project timelines, go/no-go points, and predefined goals: such strictures do not suit projects that need to take time to define the problem in context and the acceptable solutions in a mix of cultures. Findings show that language between stakeholders creates issues for ethical evaluation. Mutuality can be found, but it is within the close confines of a certain epistemic community and it does not easily reach over the social/market boundary of creating trust and mutuality with service users.

It takes time and effort to appreciate context and culture. Practice is not the world as we would wish it to be; instead, it throws up, for example, the challenges of users understanding why particular databases are chosen, why the algorithm operates in a particular way, how "an apparently unacceptable conclusion is derived by apparently acceptable reasoning from apparently acceptable premises" – Sainsbury's (1995) definition of a paradox – and, often as O'Neil (2016) found, how users view the result of AI-enabled projects.

In summary, as Figure 1 suggests and the previous literature shows, referencing context and culture is essential for ethically acceptable AI-enabled new service solutions. Instead of being informed at each design stage of how context and culture might shape the project solution for different stakeholders, the experienced complexity of AI (for collaboration) leaves service providers abstractly discussing ethical principles and guessing what users might find acceptable.

Mutuality

Following EU, the Finnish government aspires to make Finland a leader in ethical AI, and in many areas, Finland ranks highly in AI readiness. Government policy shows its drive to use AI in public service delivery, for example in the State of AI in Finland report (2020). If so, as Figure 1 suggests, we expect evidence of mutuality. We do find (from I-2, I-6, and I-10) evidence of good intentions and the seeking of the voice and contribution of users. However, in practice even putting values into action, framing the purpose, and taking a proactive approach to the context and culture are often absent – as is the stronger boundary-crossing involving power and control sharing – that we use to define mutuality.

Innovators seem to design-out mutuality instead of paying attention to decades of research on innovation in the public sector and then designing-in mutuality. This is part of the answer to our research question, to which we now turn.

6 Conclusions

Theoretical contribution

Derived from the existing theories on AI ethics (see Dignum 2019; Coeckelbergh 2020; Greene *et al.* 2019; Tännsjö 2002), trust (see Nooteboom 2002; Six 2005), and mutuality (see Yeoman 2019, Rossi 2019; Dabos and Rousseau 2004; Thibaut and Kelly 1959; Blau 1964) we build on approaches to ethical AI adaptation in public services. Our study highlights that there is a

need for a clear appreciation of all of forms that mutuality and trust will take – setting the boundaries within which any ethical evaluation will occur in AI development. In short, the answer to our research question 'what actions AI developers are taking to achieve mutuality in the deployment of AI in new public service solutions' is currently only a few.

Our results provide some possible explanations for the fact that most public organizations see the need for actions to support mutuality and trust-building but there is a struggle to include service users and an acceptance of market-dominated mutuality. We argue that mutuality and trust-building presume active engagement of users (alongside developers, funders, and service providers) in the ethical assessment. Trust and mutuality are relationships requiring active interactions. Study shows that this is challenged by missing dialogue, language barriers, centralization of power, and poor understanding of AI's purpose and AI as technology alongside political turbulence and differing approaches to complex service environments by project partners. Figure 1 provides tools for assessing mutuality and trust during each development stage to help overcome key challenges. Without such an assessment, the dangers of injustice and inequality from AI noted by O'Neil (2016) and Eubanks (2017) may be actualised.

One conclusion is that without a clear appreciation of mutuality and trust, the appropriateness of ethical principles and standards cannot be ascertained since it will be unclear if market or non-market governances (or some combination) prevail. The post-facto ethical evaluation (Jobin *et al.* 2019; Fjeld *et al.* 2020) only closes the stable door after the horse has bolted. The mutuality assessment is a pre-condition for Tännsjö's (2002) situated ethics, as is active, time-consuming user engagement (Dignum 2019).

Problem-centric AI innovation requires innovators, including service users, to analyse mutuality and trust as a basis from which to decide on ethical evaluation criteria. We reject the NPM presumption that market governances predominate; instead, mutuality boundaries should be consciously negotiated by active agents, a process our framework is designed to support. Good intentions are insufficient to provide ethical solutions: mutuality is a boundary of power and control. Mutuality can breed trusting relationships Six (2005) speaks of, founded on objective power relations that are deeper than the subjective feelings and Coeckelbergh (2020), Green *et al.'s* (2019), and Rossi's (2019) approaches to trust-building.

Implications

Our evidence suggests that top-down AI innovations in the Finnish public sector may feature injustices all concerned wish to avoid. In general terms, the absence of mutuality and trust repeats some of the failings of previous rounds of IT and digitalisation innovations. To avoid this scenario our results can help those in charge of governing AI development projects to shift viewpoint; in organising the design and delivery of AI-enabled new service solutions, learning by users, technical developers, and service providers can be distributed to achieve the compromises necessary for social, economic, and ethical acceptance. This logic-of-practice learning begins by consciously agreeing on the rules of the game – the problem to be solved and the position of the mutuality boundary to be used.

In practical terms, for public service managers and AI project initiators this research suggests (a) agreeing on the terms of mutuality in project design, and (b) the need to focus on learning and its distribution between technical developers, users, and service providers, and on how the time and attention necessary are planned for and implemented in AI projects.

However, robustness of our findings needs to be tested in further research e.g., case comparisons since contexts and cultures alter mutuality boundary and trust levels across AI projects, one limitation of this research is that this negotiation needs to be done in each context and culture: the Finnish experience cannot simply be transferred. Further research will explore mutuality in other contexts and cultures and extend from top-down projects to ecosystems without central controllers featuring a radical change in roles, relationships, and responsibilities for SLBs and middle-managers.

Acknowledgments
The data for this study were collected under the project 'Ethical AI for the Governance of the Society' (ETAIROS), funded by the Academy of Finland and the Strategic Research Council (SRC), and conducted at the School of Management, University of Tampere, in 2020-2021. The authors express their gratitude for the support received in developing their research.

Disclosure statement
No potential conflict of interest was reported by the authors.

Bibliography

AI HLEG. 2019. “Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence.” European Commission, April 2019.

Ananny, M. 2016. “Toward an ethics of algorithms: Convening, observation, probability, and timeliness.” *Science, Technology, & Human Values 41*(1), 93-117. DOI: 10.1177/0162243915606523

Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. 2019. “Transforming the communication between citizens and government through AI-guided chatbots.” *Government Information Quarterly*, *36*(2), 358-367.

Barrett, M. and Baum, D. 2017. “A model of pathways to artificial superintelligence catastrophe for risk and decision analysis.” *Journal of Experimental & Theoretical Artificial Intelligence* 29:2, 397-414, DOI: 10.1080/0952813X.2016.1186228

Bernstein, B. 2000. *Pedagogy, Symbolic Control and Identity: Theory, Research Critique*. Lanham: Rowman and Littlefield Publishers.

Bietti, E. 2020. “From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy.” In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 210-219).

Blau, P. M. 1964. *Social exchange theory*. New York, NY: Wiley.

Bryman, A. 2004. *Social Research Methods***.** Oxford: Oxford University Press.

Bryson, JJ. 2019. “The Past Decade and Future of AI’s Impact on Society, in Towards a New Enlightenment?” *A Transcendent Decade,* Vol. 11. Turner, Madrid.

Bryson, JJ., Diamanthis, ME. and Grant TD. 2017. “Of, for and the people: the legal lacuna of synthetic persons.” *Artificial Intelligence Law*, 25, 273-291.

Bryson, JJ. 2018. “Patiency is not a virtue: the design of intelligent systems and systems of ethics.” *Ethics and Information Technology*, 20, 15-26.

Bryson, JJ. and Theodorou, A. 2019. “How society can maintain human-centric artificial intelligence” in Toivonen M and Saari E (Eds), *Human-centered digitalisation and services,* Springer, Berlin.

Bryson, JJ. and Kime, PP. 2019. *Just an Artefact: Why Machines are Perceived as Moral Agents*, Proceedings of the Twenty Second International Joint Conference on Artificial Intelligence.

Bryson, JJ. 2017a. “The meaning of the EPSRC principles of robotics.” *Connection Science*, 29:2, 130-136, DOI: 10.1080/09540091.2017.1313817

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B. and Amodei, D. 2018. *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. arXiv preprint arXiv:1802.07228. https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

Castells, M. and Himanen, P. 2002. *The Information Society and the Welfare State – the Finnish Model*, OUP, Oxford.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M. and Floridi, L. 2017. “Artificial Intelligence and the ‘Good Society’: the US, EU and UK approach.” *Science, Engineering and Ethics* 27. DOI: DOI 10.1007/s11948-017-9901-7

Charmaz, K. 2000. “Grounded theory: Objectivist and constructivist methods.” *Handbook of qualitative research, 2*, 509-535. California, CA: Thousand Oaks.

Charmaz, K. 2006. *Constructing grounded theory: A practical guide through qualitative analysis*. Sage. London.

Charmaz, K. 2008. "The legacy of Anselm Strauss in constructivist grounded theory." In *Studies in symbolic interaction*. Emerald Group Publishing Limited.

Chun, A. H. W. 2008. "An AI framework for the automatic assessment of e-government forms." *AI Magazine*, *29*(1), 52-52.

Coeckelbergh, M. 2020. *AI Ethics.* Cambridge, MA: MIT Press.

Collier, M., Fu, R. and Yin L. 2017." Artifical intelligence: Healthcare's new nervous system. Edited by Accenture." Retrieved 07.01.2021 from https://www. accenture.com/t20170418T023052Z__w__/au-en/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf

Dabos, G. E., and Rousseau, D. M. 2004. "Mutuality and reciprocity in the psychological contracts of employees and employers." *Journal of applied Psychology* 89 (1): 52. DOI: 10.1037/0021-9010.89.1.5252

Daniels, H. (2016). *Vygotsky and Pedagogy*. London: Routledge.

De Bruijn, H. 2007. *Managing Performance in the Public Sector*. London: Routledge.

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.

Dignum, V. 2019. *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Switzerland, Cham: Springer Nature.

Dietz, G. 2011. "Going back to the source: why do people trust each other?" *Journal of Trust Research* 1 (2): 215–222. DOI:10.1080/21515581.2011.603514.

Dietz G. and Den Hartog DN. 2006. "Measuring trust inside organisations." *Personnel Review* 35 (5): 557-588.

Drigotas, S. M., Rusbult, C. E. and Verette, J. 1999. "Level of commitment, mutuality of commitment, and couple well-being." *Personal Relationships 6* (3): 389-409. https://doi.org/10.1111/j.1475-6811.1999.tb00199.x

Easterby-Smith M., Thorpe R. and Lowe A. 1991. *Management Research*. London: Sage.

Edwards L. and Veale M. 2018. "Enslaving the algorithm: From a "right to an explanation" to a "right to better decisions?" *IEEE Security & Privacy* 16 (3): 46-54. DOI:10.1109/MSP.2018.2701152

Etikan, I., Musa, S. A., and Alkassim, R. S. 2016. "Comparison of convenience sampling and purposive sampling." *American journal of theoretical and applied statistics* 5 (1): 1-4. DOI: 10.11648/j.ajtas.20160501.11

Eubanks V. 2017. *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor.* New York, NY: St Martin's Press.

Fina S. Heider B. Mattila M. Rautiainen P. Sihvola M-W and Vatanen K. 2021. "Unequal Finland. Regional socio-economic disparities in Finland." Europa. Friedrich-Ebert-Stiftung – Politics for Europe.

Fisher, A. 1988. *The Logic of Real Arguments.* CUP: Cambridge.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M. 2020. *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI.* Berkman Klein Centre Research Publication, (2020-1). https://dx.doi.org/10.2139/ssrn.3518482

Floridi, L. 1999. "Information ethics: On the philosophical foundation of computer ethics." *Ethics and information technology*, *1*(1), 33-52.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. 2018. "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations." *Minds and Machines*, *28*(4): 689-707. https://doi.org/10.1007/s11023-018-9482-5

Floridi L. and Taddeo M. 2016. "What is data ethics?" *Phil. Trans. R. Soc.* A 374: 20160360. http://dx.doi.org/10.1098/rsta.2016.0360

Friedman, B. and Nissenbaum, H. 1996. "Bias in computer systems." *ACM Transactions on Information Systems (TOIS)*, 14(3): 330-347. https://doi.org/10.1145/230538.230561

Gasser, U., and Almeida, V. A. 2017. "A layered model for AI governance." *IEEE Internet Computing*, *21*(6): 58-62. DOI: 10.1109/MIC.2017.4180835

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. 2021. "Attachment and trust in artificial intelligence." *Computers in Human Behavior*, *115*, 106607.

Greene, D., Hoffmann, A. L. and Stark, L. 2019. *Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning*. 10 (2019).

Gupta, KP. 2019. "Artificial intelligence for governance in India: Prioritizing the challenges using analytic hierarchy process." *International Journal of Recent Technology and Engineering*, 8 (2): 3756-3762.

Gupta, S., Kamboj, S., & Bag, S. 2021. "Role of Risks in the Development of Responsible Artificial Intelligence in the Digital Healthcare Domain." *Information Systems Frontiers*, 1-18.

Haas, PM. 1992. "Introduction: Epistemic Communities and International Policy Coordination." *International Organisation* 46 (1): 1-35.

Hagendorff, T. 2020. "The ethics of Ai ethics: An evaluation of guidelines." *Minds and Machines* 1-22. https://doi.org/10.1007/s11023-020-09517-8

Hardin R. 2002. *The crippled epistemology of extremism*, in Breton *et al*, Political Rationality and Extremism, CUP: Cambridge.

Henson, R. H. 1997. "Analysis of the concept of mutuality." *Image--the journal of nursing scholarship*, *29*(1): 77-81. DOI: 10.1111/j.1547-5069.1997.tb01144.x

Hickok, M. 2020. "Lessons learned from AI ethics principles for future actions." *AI and Ethics* 1-7. https://doi.org/10.1007/s43681-020-00008-1
Hunt, G., & Mehta, M. (Eds.). 2013. *Nanotechnology:" Risk, Ethics and Law"*. Routledge.

Jobin, A., Ienca, M., and Vayena, E. 2019. "The global landscape of AI ethics guidelines." *Nature Machine Intelligence* 1(9): 389-399.

Johnson, D. G. 1985. *Computer ethics*. NJ: Englewood Cliffs.

Jordan, J. V. and Stone Centre for Developmental Services and Studies. (1986). *The meaning of mutuality* (Vol. 23). Wellesley, Mass.: Stone Centre for Developmental Services and Studies, Wellesley College.

Kankanhalli A., Charalabidis Y. and Mellouli S. 2019. "IoT and AI for smart government: A research agenda." Government Information Quarterly 36 (2): 304-309. https://doi.org/10.1016/j.giq.2019.02.003

Kaplan, U. 2014. "Moral judgment is not based on a dichotomy between emotion and cognition: Commentary on Bazerman et al. (2011)." *Emotion Review 6*(1): 86-86.

Kaufmann, T., Gutknecht, R., Lindner, R., Schirrmeister, E., Meißner, L. and Schmoch, U. 2021.. 'Trust, trustworthiness and technology governance'. TIGTech research and consultation. TIGTech Anchor Document, Fraunhofer.

Kinder, T. 2000. "A sociotechnical approach to the innovation of a network technology in the public sector – the

introduction of smart homes in West Lothian." *European Journal of Innovation Management* 3 (2): 72 - 90. https://doi.org/10.1108/14601060010322284

Kinder, T. 2010. "e-Government service innovation in the Scottish criminal justice information system." *Financial Accountability & Management* 26 (1) 21-4. https://doi.org/10.1111/j.1468-0408.2009.00489.x

Kinder, T. 2012. "Learning, Innovation and Performance in Post-New Public Management of Locally Delivered Public Services." *Public Management Review* 14 (3): 403-428. DOI: 10.1080/14719037.2011.637408

Kinder, T., Koskimies, E. and Stenvall J. 2021. "Local public services and the ethical deployment of artificial intelligence." *Government Information Quarterly*, forthcoming.

Kinder, T. 2002. "Good practice in best practice". *Science and Public Policy*, 29:3, 1–14.

Kreps, G. L. and Neuhauser, L. 2013. "Artificial intelligence and immediacy: designing health communication to personally engage consumers and providers." *Patient education and counseling*, *92*(2), 205-210.

Kuziemski, M. and 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.

Lin, P., Abney, K. and Bekey, G. A. (Eds.). 2012. *Robot ethics: the ethical and social implications of robotics*. Intelligent Robotics and Autonomous Agents series.

Laitinen I., Stenvall J. and Kinder T. 2017. "Co-design and action learning in local public services." *Journal of Adult & Continuing Education* 1: 26, August. DOI: 10.1177/1477971417725344

Lukes S. 1974. *Power: A Radical View*. Macmillan, London.

McNamara, A., Smith, J. and Murphy-Hill, E. 2018. "Does ACM's code of ethics change ethical decision making in software development?" In *Proceedings of the 2018 26th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering* (pp. 729-733). https://doi.org/10.1145/3236024.3264833

MEAE (Ministry of Economic Affairs and Employment). 2017. *Finland's Age of Artificial Intelligence Turning Finland into a leading country in the application of artificial intelligence. Objective and recommendations for measures.* Retrieved 12.01.2021 from: *https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf*

Mikalef, P., Lemmer, K., Schaefer, C., Ylinen, M., Fjørtoft, S. O., Torvatn, H. Y., ... & Niehaves, B. 2021. "Enabling AI capabilities in government agencies: A study of determinants for European municipalities." *Government Information Quarterly*, 101596.

Mittelstadt, B., Russell, C. and Wachter, S. 2019. „Explaining explanations in AI." In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 279-288).

Moor, J. H. 1985. "What is computer ethics?." *Metaphilosophy*, *16*(4), 266-275.

Möllering, G. 2006. *Trust: reason, routine, reflexivity*, Amsterdam: Elsevier.

Nemitz, P. 2018. "Constitutional democracy and technology in the age of artificial intelligence." Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 376, 20180089. http://dx.doi.org/10.1098/rsta.2018.0089

Nietzsche, FW. 1997. *Beyond good and evil: Prelude to a philosophy of the future*. New York, NY: Mineola.

Nooteboom, B. 2002. *Trust: Forms, foundations, functions, failures and figures*. Edward Elgar Publishing.

Nooteboom, B. and Six, FE. 2003. *The Trust Process in Organisations*. Cheltenham: Edward Elgar.

O'Neil, K. 2016. *Weapons of Math Destruction: How Big Data Increases Inequalities and Threatens Democracy*. London: Penguin,

Osborne, S. P., Radnor Z., Kinder T., and Vidal I. 2015. "The Service Framework: A Public Service-dominant Approach to Sustainable Public Services." *British Journal of Management* 26 (3): 424–438. DOI: 10.1111/1467-8551.12094

Osborne, S. P. 2006. *The new public governance?* 1.

Pencheva, I., Esteve M., and Mikhaylov SJ. 2018. "Big data and AI – A transformational shift for government: So, what next for research?" *Public Policy and Administration* 35 (1): 24-44. DOI: 10.1177/0952076718780537

Radnor, Z. Osborne, SP. Kinder, T. and Baranova, P. 2014. "Operationalizing Co-Production in Public Services Delivery: The Contribution of Service Blueprinting." *Public Management Review* 16 (3): 402-423. DOI: 10.1080/14719037.2013.848923

Rahwan, I. 2018. "Society-in-the-loop: programming the algorithmic social contract." *Ethics and Information Technology* 20 (1): 5-14.

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., and Floridi, L. 2020. "The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation." *AI & SOCIETY*, 1-19. https://doi.org/10.1007/s00146-020-00992-2

Roberts, M. 2016. *The Long Depression – How it Happened, Why it Happened and What Happens Next.* Chicago, IL: Haymarket Book.

Rosenberg, N. 1982. *Inside the Black Box - Technology and Economics.* New York, NY:Cambridge University Press.

Rossi, F. 2018. "Building trust in artificial intelligence". *Journal of international affairs*, *72*(1), 127-134.

Rossi, P., and Tuurnas, S. 2021. "Conflicts fostering understanding of value co-creation and service systems transformation in complex public service systems." *Public Management Review*, *23*(2), 254-275.

Ryan, R.M and Deci, E.L. 2000. "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being." *American Psychologist* 55 (1): 68-77. DOI: 10.1037110003-066X.55.1.68

SAIP. 2019. "Leading the way into the era of artificial intelligence: Final report of Finland's Artificial Intelligence Programme 2019." Steering group and secretariat of the Artificial Intelligence Program. Publications of the Ministry of Economic Affairs and Employment 2019:41 Helsinki.

Sainsbury, RM. 1995. *Paradoxes*. CUP: Cambridge.

Schaefer, C., Lemmer, K., Samy Kret, K., Ylinen, M., Mikalef, P., & Niehaves, B. 2021). "Truth or Dare?–How can we Influence the Adoption of Artificial Intelligence in Municipalities?." *In Proceedings of the 54th Hawaii International Conference on System Sciences* (p. 2347).

Schattschneider, EE. 1975. *The Semisovereign People: A realist's view of democracy in America*. Illinois: Dryden

Schummer, J., & Baird, D. (Eds.). 2006. *Nanotechnology challenges: implications for philosophy, ethics and society*. World Scientific.

Silverman, D. 2001. *Interpreting qualitative data: Methods for analysing talk, text and interaction* (2nd edition). Thousand Oaks, CA: Sage.

Six, FE. 2005. *The Trouble with Trust – The Dynamics of Interpersonal Trust Building*. Cheltenham: Edward Elgar

Spieth, P., Schneckenberg, D. and Ricart JE. 2014 "Business model innovation–state of the art and future challenges for the field." *R&D Management* 44 (3) 237-247. https://doi.org/10.1111/radm.12071

Stinchcombe, AL. 1990. *Information and Organisations*. California, LA: University of California Press,.

State of AI in Finland. 2020. Retrieved 07.01.2021 from https://faia.fi/market-research/

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., ... and Teller, A. 2016. Artificial Intelligence and Life in 2030. One-hundred-year study on artificial intelligence: Report of the 2015-2016 Study Panel. *Stanford University, Stanford, CA, http://ai100. Stanford. Edu/2016-report*. Retrieved 03.10.2019 from https://ai100.stanford.edu/sites/default /files/ai_100_report_0831fnl.pdf

Thibaut, J. W. and Kelley, H. H. 1959. *The Social Psychology of Groups.* New York, NY: Wiley

Trocin, C., Mikalef, P., Papamitsiou, Z., & Conboy, K. 2021. "Responsible AI for digital health: a synthesis and a research agenda." *Information Systems Frontiers*, 1-19.

Tronick, E. D., Als, H., and Brazelton, T. B. 1977. "Mutuality in mother-infant interaction." *Journal of communication* 27 (2): 74-79.

Tännsjö T. 2002. *Understanding Ethics.* Edinburgh University Press, Edinburgh.

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., and Wanjiku, W. G. 2020. "Framing governance for a contested emerging technology: insights from AI policy." *Policy and Society* 1-20. DOI: 10.1080/14494035.2020.1855800

Vygotsky, L.S. (1934). *The collected works of LS Vygotsky: Vol. 1, Problems of general psychology*, Rieber RW & Carton AS (Eds). New York, NY: Plenum.

Wartofsky, M. 1979. *Models, Representation and Scientific Understanding*. Reidel, Dordrecht.

Watson, H. J. 2014. "Tutorial: Big Data Analytics: Concepts, Technologies, and Applications," *Communications of the Association for Information Systems* 34 (65). http://aisel.aisnet.org/cais/vol34/iss1/65

Weber, M. 1978. *Economy and Society: An Outline of Interpretive Sociology*. CA: University of California Press.

Weibel, A., Den Hartog, DN., Gillespie, N. Searle, R. Six, FE. and Skinner D. 2016. "How do controls impact employee trust in the employer?" *Human Resource Management* 55 (3): 437-462. DOI:10.1002/hrm.21733

Weibel, A. and Six, FE. 2013. "Trust and control: the role of intrinsic motivation," in Bachmann R and Zaheer A (Eds.*), Handbook of advances in trust research* (pp. 57-81). Cheltenham: Edward Elgar.

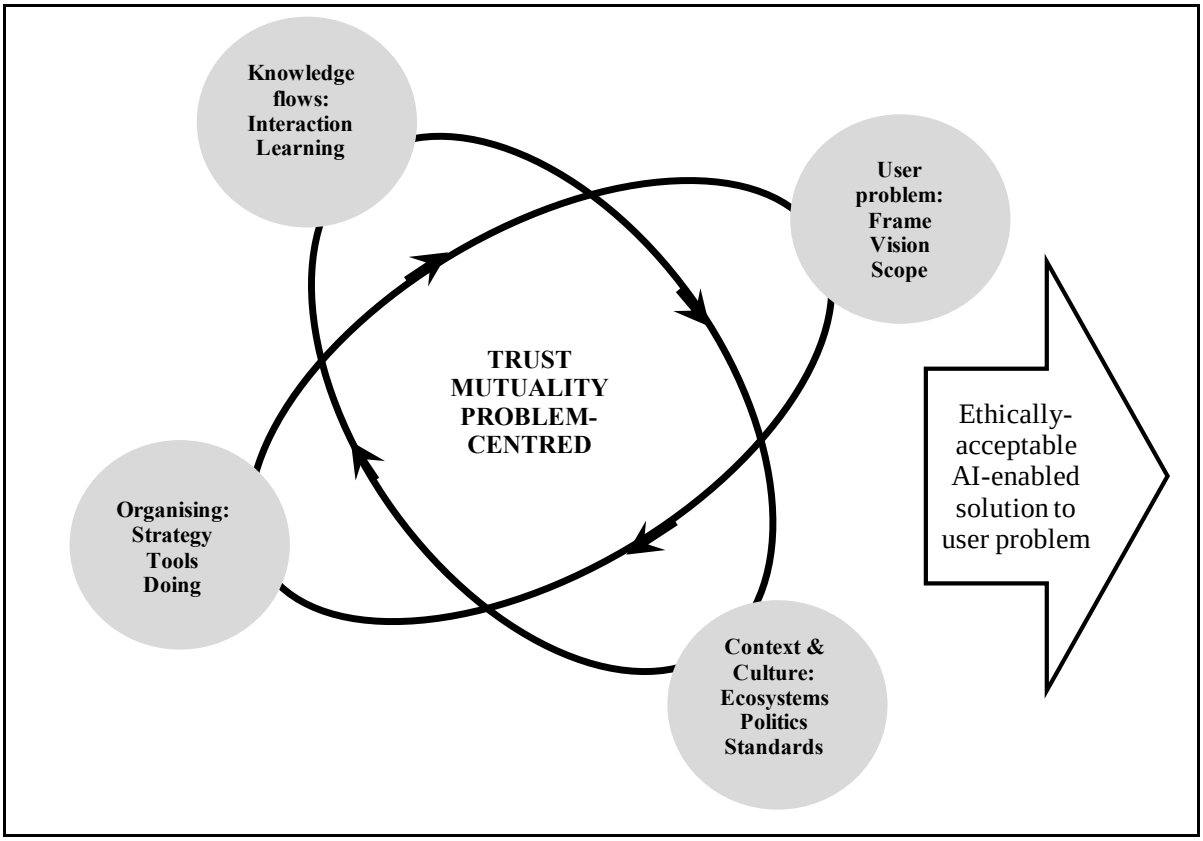
Weick, K.E. 1995. *Sensemaking in organizations,* (Vol. 3). CA: Sage, Thousand Oaks,

Winfield, A. F., & Jirotka, M. 2018. "Ethical governance is essential to building trust in robotics and artificial intelligence systems." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376 (2133): 20180085. http://dx.doi.org/10.1098/rsta.2018.0085

Wirtz, B. W., and Müller, W. M. 2019. "An integrated artificial intelligence framework for public management." *Public Management Review 21* (7): 1076-1100. DOI: 10.1080/14719037.2018.1549268

Yeoman, R. 2019. *Ethics, Meaningfulness, and Mutuality*. New York, NY: Routledge.

Yin, RK. 2014. *Case Study Research Design and Methods* (5th ed.), Thousand Oaks, CA: Sage



Key actors		
I-1	Male	Digital population and data service agency
I-2	Male	Ministry of Finance
I-3	Male	Finnish Center for Artificial Intelligence (FCAI)
I-4	Male	Technology and software provider
I-5	Male	Prime Minister's Office
I-6	Female	The Finnish Innovation Fund SITRA
I-7	Male	Finance Finland (interest organisation)
I-8	Male	Ministry of Finance
I-9	Male	Aalto University / Nokia Bell Labs
I-10	Female	Ministry of Finance

Basic statement	Initial code	Focused code	Theoretical coding
"But one conflict comes when technographics develop AI. So, professionals and regulators and officials are the ones that think for the people." (I-6)	Lack of user engagement The value of citizen involvement Dialogue Technology push	Knowledge flows	Social mutuality framework

Figures: