Appendix S2 Data Evidence Linking First-Order Codes, Second-Order Themes and Agregate Dimensions

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
G1	(P1.1) Shift attention from explaining AI to explaining the decision supported by AI. (P2.5) Explanation of the rationale of individual decision-making (P4.4c) Provide meaningful explanations about AI decisions (P6.1) Improve and optimize the quality and clarity of explanation of AI decisions (P11.8) Care managers could not explain why particular types of care services were recommended (P17.7) There is a continuous development of the texts explaining the motivations for all decisions, positive and negative (P18.4) AI can increase the chance of administrative evil due to its inscrutability. AI lacks explainability or requires technical expertise to understand decisions (P20.1d) Also, we want to be able to explain decision support, so that's why we need explainability in our model and information chain.	Explainable decision (Explaining the decisions)	G1; G2; G3: **Explainability**
G2	(P1.3) From explainable algorithms to explainable processes (P9.6a) the basis for debt calculations and the debt-recollection process not clearly explained (P10.17) AuroraAI is largely a black box whose recommendations and predictions are not transparent and remain difficult to explain (P14.7) Explainability is a key challenge for AI-based decision-making. How did an AI reach a decision with the input data it received? (P17.2) Our responsibility is to be very clear about the decisions and how the robot makes them (P20.8) If more advanced algorithms (inscrutable models) outperform explainable ones, we use the more advanced ones and focus on how to make them explainable (P21.9) The teachers sued the district through their union, arguing	Explainable process (Explaining the process)	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	that the software was fundamentally inscrutable and that there was no way for teachers to know whether the software was accurately assessing their job performance (P24.2) How certain patterns and outputs are generated by algorithms has been referred to as an algorithmic 'black box', as these algorithms are not readily understandable to humans, making them unexplainable to citizens (P26.14) Algorithmic decisions can be explained at the level of the model (e.g., how does it function? What assumptions does it make?) or at the individual record level (why was this record classified as A rather than B?)		
G3	(P20.9) The quest for explainable AI is made more complex by the diversity of explanation-related requirements among various internal and external DBA stakeholders. Each of them requires a specific kind of explanation of a given model's internal logic and outputs (P24.7) A lack of explainable models and outcomes is at the heart of the debate on the algorithmic black box. Algorithmic explainability takes different forms and the type of explanation needed depends on the user of the algorithm	Stakeholder-based explanation (Providing meaningful explanation to different stakeholders)	
G4	(P4.4b) Being transparent about how and when using AI (P5.9a) Algorithms will only be sufficiently transparent if government creates and maintains records that document their objectives for algorithms and vendors disclose sufficient information describing how algorithms are developed (P6.2) Clarity on the decision-rules (algorithms) and data used for a specific decision. / Provide insights into decision components, calculations applied, etc. (P7.3) Lack of transparency as no explanation was provided on the variables, weighting and correlation used by COMPAS (P8.1) Provide information about how decisions are reached, the	Process transparency	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	process involved and the data used. / Recognise the right to explanation on the logic involved. (P13.8) Additionally, process opacity is another transparency challenge. It is unclear how exactly algorithms are used by policymakers. (P14.5) ML algorithms add another layer of opaqueness to their decision-making processes (P19.2a) Public authorities implementing ADM systems are often constrained by suppliers' confidentiality clauses, considered a major impediment to ADM-related transparency and therefore accountability. Public procurement practices, including relevant contractual standards are thus considered potential focus to reduce ADM-related harms (P19.5) Although 'black box' AI opacity was considered relevant, the focus here was on 'process transparency' (processes surrounding design, development and deployment) more than 'system transparency' (operational logic) (P20.1c) The agency can be taken to court when we end a company by means of the law. In that situation, in court, we have to provide full documentation of why that decision has been made (P25.2) The opaqueness of AI technology is accepted in the private sector, but it challenges government transparency. User will rightly have quite different expectations of their right to understand how decisions on their benefit entitlement or health care coverage have been made (P26.6) Records were selected based on two criteria: "the documentation of what government did, why and how" and "the value of the records for future historical research" .Departments create and publish appraisal policies outlining how decisions are undertaken and what material will be selected (P26.13) To understand the archive fully you must understand the processes through which "the selection and preservation of 'valuable' information occurs" (P26.17) The processes and rationale behind training data curation should be published to help identify potential biases		G4; G5; G6; G7: **Transparency**

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
G5	(P2.2) Model transparency requires concerted efforts of developers, computer science community, the industry and the public sector (P8.2) Share information on algorithms in a way that can be clearly understood by the public (P13.7) Technology-related forms of opacity (internal and external opacity) contribute largely to the lack of transparency of algorithms. (P24.1) The algorithm's parameters and decisions were never published and investigated residents were not informed they were investigated for welfare fraud. The system was prohibited for a lack of transparency of its algorithm (P21.7) Because these tools are often based on private systems licensed to government agencies, the design specifications and the inner workings are considered trade-secrets and are not publicly available (P22.2) People affected by ADM do not generally have the opportunity or capacity to review data, codes etc. Not only are these things difficult for the nonprogrammer to understand, companies often go to great lengths to protect their code as trade secrets (P24.9) Transparent algorithms improve trustworthiness among the general public (P2.6)) It is recommended that models are a priori well understood, that their predictive features are understandable (P2.1) Interpretable and comprehensible model (P10.3) AI introduces another layer of complexity making it difficult for even experts to understand its inner workings (P20.7) I'm very fond of transparency. I think it's the way to go that it's fully disclosed why a system reacts [the way] it does	Algorithm (model) transparency	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
G6	(P24.3) Algorithms are sometimes deliberately made inaccessible as they are often developed by commercial parties and subject and protected by intellectual property (P24.4) Inaccessible algorithms can perpetuate the generation of decisions or recommendations that are biased, discriminatory, or inaccurate (P24.5) Accessibility goes beyond public availability, but also that external independent experts can access an algorithm for inspection and analysis to assess if it is compliant and does not violate any rules (P24.6) Accessibility does not only concern the algorithm, but also the underlying data. The data must be accessible to independent experts	Accessibility Two forms (i) algorithm accessibility: transparency-as-code-availability and transparency-as-code auditability and (ii) data accessibility	
G7	(P4.5a) Being open to sharing source code, data and other relevant information (P7.4a) To tackle "the legal black box", relevant laws must be amended to compel disclosure to court-approved parties or expert committees (P11.2) Information communicated to the public about AI use matters to initial trust in the services (P19.1a) For ordinary people, achieving basic awareness (disclosure of the existence of a relevant system, notification, knowing what's happening, overcoming secrecy) (P19.3) At a minimum, organizations should notify people that relevant algorithms are being used. Lack of public information on ADM is a major driver of legal contestation	Self and regulated disclosure	
G8	(P2.4c) Scrutiny of unanticipated failures by domain experts and citizens (P2.8) Regulatory efforts (e.g. model certification, impact assessments) are thus vitally needed to ensure that AI tools are brought to bear in a thoughtful and effective manner to bolster algorithmic accountability.	Regulatory/public accountability	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	(P10.5) Public sector AI use needs to be transparent and comply with the need for regular scrutiny and oversight (P14.6) Identifying who (or what) has responsibility or liability for an AI decision is challenging (P19.1b) Questioning decision-makers' presentation of systems, revisiting previous authority based on incomplete understanding (P19.7) The idea that government may be not only failing to regulate ADM but even exploiting internal opacities to evade scrutiny is at the heart of one of the most recent and contentious cases (P22.7)) Technologies must be accessible to outside scrutiny especially by those most affected by their risks (P25.3) Need to design appropriate accountability frameworks to prevent politicians and policy-makers from taking advantage of blame-shifting		G8; G9; G10; G11: **Accountability**
G9	(P17.9) The caseworkers understood that the ADM had relieved them of some decision responsibility, but were still concerned with their own accountability (P19.2b) Lack of clear responsibility described in the starkest terms in the Criminal Justice group as 'agency laundering', or delegation of discretionary authority to technology	Internal accountability	
G10	(P13.2) Account for Algorithms to multiple critical audiences (e.g. technical and non-technical audiences)	Multiple critical-audience accountability	
G11	(P21.1) Like other govt.-commissioned private actors, developers of AI systems used in government decision-making should be treated as state actors for purposes of constitutional liability using the state action doctrine. This is a necessary step to bridge the current AI accountability gap (P21.2) To prevent government agencies from using different rules	Constitutional accountability	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	or means that allow AI systems to be used as accountability-avoidance mechanisms when companies cause constitutional violations, the state action doctrine remains a powerful and flexible common law approach to redress this gap (P21.3) In many of the instances, employees could not provide remedy for the harms caused by the proprietary AI at the heart of the constitutional liability concerns. Unless vendors are subject to the court's jurisdiction, it cannot assert any oversight or impose any specific injunctive relief (P21.4) Treating AI vendors as state actors would allow those who have been constitutionally harmed to sue the vendors directly and give courts access all relevant information about the AI system, its function, and the role of the vendor in the alleged constitutional violation (P21.8) In all the cases throughout the country, constitutional accountability for the creators of the AI systems responsible for the harms has been entirely absent (P22.4) Accountability gaps arise when governments circumvent checks on their power by outsourcing authority to private companies or when public actors are not able to provide remedies for harms caused. State action doctrine can be deployed to ensure that accountability for harms is not evaded		
G12	(P1.4) Shift from an instrumental to an institutional approach by establishing countervailing structures (P2.4a) Proper and independent system vetting and continuous monitoring of their functioning (P4.1) Scrutinize automated decision-making systems at the stage of goal-setting, procurement and implementation (P21.5) Concerns have been raised over the use of AI systems often deployed without adequate assessment, safeguards, [or] oversight. (P23.3) Given proprietary concerns, it is advantageous to establish	Institutional (regulatory) governance	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	industry-wide standards and a level of government involvement in the certification that these standards are being met		
G13	(P2.3b) Test technical and governance implications of models – broader considerations of fairness, bias, and transparency - (collaborative model testing and validation by developers and domain experts) (P5.8) AI governance structures must be populated by the widest range of stakeholders including policy makers, government domain experts, and AI systems developers inside and external to government, along with individuals who will be affected by AI decision making to address questions relevant to trustworthiness (P7.5) Ensuring the ethics of designers and holding them accountable require multi-stakeholder deliberation. / A technologically-informed and socially-apt governance model is only fruitful via constructive multi-stakeholder dialogue (P9.13) Implementing organizations can adopt best-practice frameworks to ensure ethical use of AI (P10.4) Constant ethical self-examination, vigilance and deliberation by ethical experts, scientists, technology developers, and other relevant stakeholders alongside AI development (P10.9) The Ethics Board followed a proactive ethical deliberation process entailing anticipation, involvement and multidisciplinary expertise (P10.21) Ethics board structure should be running before AI programme starts (P14.10) Another approach to AI governance is via an independent quality assurance mechanism to test its compliance with ethical/legal considerations (P18.8) Because AI deals in quantification and numerical representation and feature engineering, it may increase the chance of administrative evil through masking and moral inversion (P19.4) Broader risk assessment and mitigation strategies around ADM initiatives were considered desirable, including	Multi-stakeholder ethics governance	G12; G13; G14; G15; G16: **Ethics and Governance**

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	consideration of legal as well as ethical frameworks and paying special attention to stakeholder engagement		
G14	(P4.4a) Measure the impact of using AI (P5.9b) Another strategy likely to improve trustworthy AI development is algorithmic audit (e.g. via impact assessments) (P8.4) Carry out pre-and post-implementation impact assessment of algorithms (P8.5) Carry out impact assessment of the damage that access to source code may entail in contexts identified by law (P9.8) Little testing/piloting was conducted. / No modelling was done (P13.4) Testing the impact (advantages and adverse effects) of algorithm in real-world setting is key to demonstrating its reliability. / This can be achieved via field experiments (P14.11) Need for practical tools and processes (e.g. impact assessment.) to assess AI for safe and ethical use (P23.4) Certification testing of AI systems can be carried out to identify design defects (by using AI itself). This can also speed up the process of ethical audit so it does not delay deployment (P5.7) Experimenting with low-risk application contexts (P16.4b) It is important to assess whether the rules implemented within the algorithms are ethical (P18.7) AI increases the chance of administrative evil because it may be enthusiastically deployed without appropriate testing or without assessing potential risks or when it might not be the optimal available solution (not fit for a given task or context) (P22.5) Risk assessment must be integrated into systems of technological development and design, to address risks to human rights before they are locked in, and perhaps, made a requirement for private companies (P13.1) To stimulate creation of sigma-type values and minimize the occurrence of adverse effects on theta-and lambda-type values,	Risk and Impact Assessment	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	assessment of the quality of the information the algorithm provides is necessary		
G15	(P16.4a) It is important to assess whether the entire system is politically legitimate	Political Legitimacy	
G16	(P23.1) One approach is to make ethics an internal algorithm design criterion from the start and attribute value to it (e.g. make it a requirement-to-buy or a weighted factor in competitive source selection). Additionally, government should fund research into ethics-by-algorithms	Ethics-led design	
G17	(P3.3) A human decision maker is perceived as more trustworthy in carrying out a decision in terms of competence, honesty, and benevolence (P9.2a) The redesigned process shifted decision-making from humans to an ADM artefact resulting in limited human agency (P11.9) Service users and their families wish to be received by a human and not a machine. / They did not embrace AI and were resistant to the new technology (P9.12) Minimizing human agency and maximizing algorithmic agency can create unintended outcomes as demonstrated in the Robodebt's case	Human agency	
G18	(P1.7) From algorithms replacing professional decision-making to professionals challenging algorithmic decision-making by applying their tacit knowledge (P7.4c) Discretionary processes must require human intervention. / ADM must not restrain the decision-maker in exercising discretion (P12.2) Can AI be entrusted to recommend a decision that is compliant with the law? / This depends on the degree of discretionary power conferred on the administrative agency (P12.3) In discretionary decision-making tasks, the need to take	Human discretion and intervention	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	unstructured information into account and determine applicable norms require a full understanding of facts and complexity of a case at stake (P12.4) Additional complexity occurs when an automated system has to take not only a data-driven but a values-driven decision. / Striking a balance between competing rights and interests (P12.5) Applying uniform rules and criteria across all decisions without considering the specificity of each case may result in unlawful decisions (P14.4) Automating administrative decision making has reduced human discretion, with concerns that an individual's circumstances may be overlooked (P16.1) Actors interviewed emphasized the discretion of humans and the non-binding role of the risk scores generated from the ADM (P16.2) Though COMPAS played a central role in decision-making, other sources of information were consulted (P16.3) In contexts requiring human discretion, algorithms can either provide evidence to strengthen a decision or form an anchor point from which a human only rarely deviates (P18.3) AI may increase the risk of administrative evil when it limits the discretion of agents disposed to prevent or resist evil (P20.1b) Internal users at the DBA can overrule the models if necessary. The human always has the power to ignore the models' recommendations if they seem questionable. A decision suggested by AI model is always verified by a case worker		G17; G18: G19**: Human agency and oversight**
G19	(P3.2) Use of AI should be controlled based on the task complexity (P5.6) Substantial human oversight and domain expertise would seem to be minimal requirements for AI system development (P3.5) Determine the degrees of automation and human accountability to be integrated into AI applications	Human oversight and control	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	(P9.2b) Minimized human oversight resulted in full automation of debt-collection, putting ADM at the centre without human scrutiny (P9.2c) Onus of proof of debts transferred to citizens. / Limited employee-citizen interaction even where help could have been provided (P9.2d) Mindful human involvement removed from process. / ADM artefact became main decision-maker rather than a support tool (P10.2) The most demanding goal is to understand the system and gain the knowledge to control its performance (P11.1) The assurance that "humans are still in the decision loop" was important to respondents. / Reveals reservations about complete AI takeover in care planning (P11.3) The public might not feel comfortable with AI replacing humans in decisions making and want humans in the loop as a safeguard (P11.4) Given the many concerns on AI, the public is keen to know that some functions are still reserved for humans (P11.10) Concerned individuals are not ready to see care planning handled completely by AI (P12.1) Need for human control with decision-makers given sufficient authority to control system and deal with undesired outcomes (P12.6) In complex government decision-making cases where a lot of variables are at play, increasing human control is required with human actors having a role "in the loop" (P12.7) Some degree of human control, intervention and discretion is necessary to avoid use of non-representative, inaccurate data, etc. (P12.8) Humans remain accountable for AI systems and given their deficiencies, there is need for meaningful human control (P17.1) Final decision on applications is made jointly by a caseworker and the technology (P17.4) Though the ADM's decisions are based on rules, humans are still needed to resolve complicated negative decisions		

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	(P17.10a) Caseworkers still needed for the review of complex negative decisions and for complicated applications (P17.10b) Even where ADM replaces human, experts are still needed to oversee the use and development of the automation (P17.10c) Although caseworkers still have the final responsibility for decisions, RPA has an increasing role in daily practice (P17.11) A hybrid model involving humans and technology can be used for core social assistance decisions (P18.1) A future role for individual bureaucrats might be to audit and oversee the operation of AI systems (P18.2) AI increases the chance of administrative evil by introducing automation bias – tendency to be overly-reliant on ADM tool without critically reviewing its decisions (P20.1a) To control AI systems' abilities and limitations, development was broken into incremental stages of multiple small-scale solutions to limit malfunctions and curb failures from escalating. Human stakeholders are tasked with judging the correctness of the model's operation, deciding whether to proceed with certain rules or not (P20.1e) Expert case workers are allowed to set thresholds for the model to make certain it produces the most useful and precise recommendations. Some guidance thresholds are set by us which we are free to move up and down (P20.2a) To avoid undesired outcomes, the organization maintained full control over the learning process by not using online-learning models that continuously learn autonomously from incoming data. Meaning that we train a model to a certain level and then we accept that it will not become smart until we retrain it (P20.3b) When not sure if the model is right or wrong, cases are pushed to the case workers. Recommendations generated by the AI are guided by case workers who rely on organizational objectives and legislative limitations. Final decisions are produced at the intersection of actions by humans and AI (P20.5) AI's mindless and error-prone nature necessitates careful control of the AI's agency and autonomy in the implementation.		

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	Humans can serve as important counterweights in this equation (P22.3) Oversight ability is crucial when automated systems are making life-and-death decisions (P22.6) Without appropriate safeguards, deployment of AI for social good can mask root causes and potential risks on vulnerable people (P23.2) An additional approach is to focus on the process of algorithm design rather than the algorithm itself, adopting measures like dual or multiple checkers with equal qualifications and redundant safeguards used in complex systems. Such a design process control will reduce the likelihood of errors with AI applications (P26.1) It is perceived that any use of ML would aim to reduce the manual burden on RMs, still allowing them to retain the final decision on permanent preservation (P26.3) The newly loaded documents are given suggested labels by the ML classifier but they are not final until they have been approved or corrected by the RM (P26.16) Record selection requires knowledge of the collections and events in the outside world. Thus, technology is expected to make the jobs of experts possible, rather than relying on it to perform the selection task for them. Maintaining human control is key in the process		
G20	(P1.2) An approach of co-creation with the public and interested parties can be taken, resulting in negotiated algorithms in which every stakeholder has its say, and a consensus needs to be reached (P10.1) Sustainable development and deployment of AI require dialogue and deliberation between developers, decision makers, deployers, end-users, and the public (P20.4) The interviewees never considered a purely technical solution for limiting AI agents' capabilities. Rather, such actions were carried out via iterative negotiations that took into account several stakeholder views, responsibility to society, and particular	Negotiated design process	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	implications for the personnel's work processes (P3.4) Develop better socio-technical understanding of applications—including how perceptions of citizens are affected—to integrate those considerations into the design process		
G21	(P5.9c) It is essential that a wide variety of AI stakeholders be involved in the processes of participatory AI development and testing in contestable contexts (P9.7) Employees' complaints prior to implementation were dismissed and the agency did not involve other relevant stakeholders in the design of the system (P10.6) Sustained public engagement is required in the design, development and deployment of AI to ensure a trustworthy process (P10.20) Governments should foster and facilitate societal discourse on the desirability of AI, and include active participation of various stakeholders and citizens (P15.3) Early engagement is necessary for clarification and mutual learning, allowing the solution to be grounded in shared meanings, with stakeholders having a voice in the process (P20.6) Organizations deploying AI requires a clear implementation strategy that takes into account the wide spectrum of stakeholders. Removing stakeholders from the process of designing, implementing, and using it will increase likelihood of failure. (P22.1) AI design should incorporate the voices of the marginalized people likely to suffer most. Participatory design also means working with practitioners to ensure their needs and those of human rights defenders for privacy, security, and protection from harm are met	Participatory design/Co-design	G20; G21;G22: **Stakeholder-based design**

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
G22	(P2.3a) Critical and ongoing cooperation between system designers and domain experts (P15.2) The lack of early problem definition, also means that caseworkers - who directly interact with both the systems and the citizens - are often left out of fundamental design decisions (P26.4) Although the solution requires technical skill to use it, the problem and resulting ML pipeline were modelled from a RMs perspective (P26.18) Engage with suppliers in the early stages of product development to influence their future development so that they work for RMs and archivists	End-user (public sector staff) engagement	
G23	(P2.4b) Adequate training of public sector staff (P4.5c) Provide sufficient training for government employees developing and using AI (P9.10a) Lack of information and training for staff prior to system go-live (P9.14) Governments need to develop sufficient capacity and competency to run ADM programmes in a responsible manner (P10.8) Skills challenge including limited machine learning knowledge plagues delivery of AI services in the public sector (P25.4) Governments could prioritize the development of expertise and capacity in AI to foster innovation and overcome some of the recurring challenges (P26.15) There is also a need for an educational programme for non-technical staff to understand the concepts of ML and their role in creating the data to train the system	AI Literacy for the public sector	G23; G24; G25: **AI literacy**

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
G24	(P5.3b) Data literacy as an imperative for government employees involved in data collection, manipulation and use (P26.9)) Records management teams should be equipped with tools for data mining and analysis, and the requisite skills to use them	Data Literacy for the public sector	
G25	(P8.3) Recipients must possess necessary skills to understand and analyse algorithms (P10.16) The Ethics Board noted the importance of citizen digital education as a requirement to bridge the digital divide	Citizen AI literacy/digital education	
G26	(P1.6) From algorithms to value-sensitive algorithm design (P2.7) Striking a balance between different value-trade-offs (P3.1) Give more attention to the balance between instrumental and value-based qualities (P13.5) Consideration must be given to how taboo tradeoffs between "secular values" – e.g. efficiency gains and "sacred values" – e.g. discrimination, is addressed and answers around this must be formulated with the citizens (P13.9) Managing the adoption and implementation of algorithms entail tradeoffs between values	Balancing multiple values	G26**: Value tradeoffs**
G27	(P4.2) Is AI facilitating the power shift between governments and citizens or intensifying existing distribution? (P10.18) Questions like What tools will the user have to be able to track the use of their data? / Can a user block a particular service provider? / Can a user see which organizations have their data? (P7.4b) The "technical black box" issue could be addressed with some opt-in or opt-out mechanisms, allowing some degree of autonomy to those subject to ADM (P19.2c) Relationships with clear power imbalances. e.g. users having no real choice but to accept terms imposed by systems or in	Addressing Government-Citizen power imbalance	G27**: Choice and autonomy**

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	terms of citizen surveillance		
G28	(P15.1) Implementing AI solutions in politicised contexts should start with problem formulations (via ethnographic studies) including end-users' roles in data-driven public services (P15.4) If we do not problematize the politics inherent in the technological solutions we build, we may prevent them from having their full effect in practice (P15.5) Problematization is a meaningful process for everyone involved in the design of technologies	Problem definition and formulation	G28; G29: **Problem and Goal clarity**
G29	(P4.3)What goals should be pursued? / Whose benefits should be prioritized? What values should be supported? (P9.1, P9.1a, P9.1b) Top management limited vision. / ADM implementation driven by economic imperative resulting in tunnel vision (P9.2) Top management limited vision resulted in an ADM with limited human agency, devoid of best practices (P10.10) A major strategic question is what the focus of AuroraAI should be; macroeconomic or highly personalized? (P10.11) It was difficult to understand the overall objective and emphasis of the AuroraAI programme. / Unclear vision. (P11.5) Another factor in building user trust is for the agency to communicate the purpose and benefits of introducing the AI to the public (P11.6) Intentions of the public agency for using AI can be a basis of public trust. / Have to be citizen-centred rather than benefiting the service provider (P18.5) AI can increase the chance of administrative evil by increasing organizational goal displacement / Organizational Value Misalignment	Goal setting, benefits and value prioritization	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
G30	(P4.4d) Offer opportunities to review and challenge decisions (P7.2) Ensure the defendant has the means and right to challenge the algorithm before the court (P9.9b) Challenging the debts proved difficult and stressful (P9.11) Due to insufficient legal support mechanisms many citizens received little help with challenging the ADM decisions (P14.9) ADM requires legal clarification around responsibility, appeal and redress (P17.5a) Citizens have the right to appeal negative decisions (P17.5b) A citizen can file an appeal against a negative decisions and assistance with the review of the appeal is provided by the Agency (P19.1c) Realising the potential to contest outcomes	Legal contestation and recourse Recourse mechanism	G30**: Recourse**
G31	(P19.6) Law and legal processes offer legitimate channels for public participation in efforts to implement ADM. Legal contestation could be an integral part of the trust dynamics. Developing means for ex ante / ongoing, as well as ex post legal inputs to AI-related decision processes are important to high quality technology implementation	Constructive use of legal contestation processes as input for policy feedback	G31: **Policy Learning**
G32	(P4.5b) Protect personal information and national security (P5.2)) Safeguard access to personally identifiable data (P5.4) Address the issue of algorithm-related privacy problems (P10.7) Public users are worried about novel challenges to privacy and security of data (P10.12)) Need to use personal data responsibly and ethically without compromising privacy and autonomy (P10.13) Data privacy issues that come with personalizing services must be taken seriously (P10.19) Ensuring the anonymity of users should be an essential part of the ethical use of AuroraAI (P14.8) AI development should incorporate ethical and legal considerations regarding data protection and privacy right from the	Manage privacy issues, anonymity, and data security	G32: **Privacy, protection and security**

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	start (P23.6) AI is so thoroughly penetrating that individual identity can almost always be unwound. Technical approaches to enhance privacy and true anonymity in databases used by AI should be adopted without spoiling the data's use		
G33	(P5.1) Problematic data introduce risks that can lead to economic devastation for industry and that may erode trust in and the legitimacy of government. Thus, trustworthy AI rests on the quality of this fuel (P5.3a) Implement enterprise data management as an essential foundation for trustworthy AI/. Data management is relevant at two points in the system (P9.3) Incompatible data sources led to error-laden decisions and bias by the ADM (P9.4)The two data sources provided to the OCI algorithm were inconsistent (P9.5) Inability of algorithm to account for citizens' unique circumstances in calculating debts. / Inability of algorithm to satisfy requisite variety (P10.14) It remains unclear how the cluster data will be collected, what the data update frequency would be and how a cluster's profile would function for people outside the cluster (P13.6) Algorithms are linked to increase in unfair outcomes and biases, which is mostly due to the already biased underlying data they are trained on (P14.1) The problematic outcomes of AI can result from the training data (e.g. incomplete, inaccurate data or data reflecting historical structural inequalities)	Data quality and context	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	(P18.6) AI increases the chance of administrative evil by reducing the amount or quality of data used for a decision: AI reinforces the primacy of quantitative data (quantification bias), even where they ("proxy variables") are less representative of the phenomenon of interest (P20.2) The crucial importance of the data used in AI systems' training is widely acknowledged. Close control of the training data is important (P20.2b) Paying attention to training data stimulates internal discussion of the data's suitability and of possible improvements in detecting problematic cases that are flagged for manual processing (P20.2d) We had a case years ago where there were a lot of bakeries that did a lot of fraud, but now it doesn't make sense to look for bakeries anymore, because now these bakeries … are selling flowers or making computers or something different (P20.3a) The selection of input sources is thus closely tied to conceptions of data quality. Our main problem was that the input data was [of] very varied quality. Return of bad results reflects the low quality of the input data (P21.6) The constitutional liability problems were exacerbated by the fact "many states simply pick an assessment tool used by another state, trained on that other state's historical data, and then apply it to the new population…" (P23.5) The next thing to tackle in ethical AI is the data the algorithm is trained on, being that data sets come from a wide variety of huge caches (P25.1) Studies have noted the importance of the data environment in predictive risk scoring, suggesting such systems are extremely vulnerable to bias, especially where data are derived from the criminal justice system (P25.6) Many of the causes of bias and unfairness in machine learning come from the training data (P26.2) The dataset was highly curated and belonged to one relatively small government department. Good performance		G33; G34: **Data Management**

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	against this dataset does not imply that equal performance would be achieved against the documents of a larger, more complex department (P26.5) A weighted approach to splitting data was used to ensure every class was represented in the training data. The danger of not doing this is that if a class does not appear in the training data it will be unknown to the algorithm (P26.7) It is clear from what we have seen that the quality of the training data is critical to good results. Supplier B reported that selection of training data had the biggest influence on accuracy (P26.8) ML pipelines could be enhanced by improving data collection by recording the "process of data collection" and relying on multilayered, and multi-person systems rather than "a single ML engineer" when compiling a dataset (P26.10) In real-life applications data often comes disorganised and are always a mixture of file types, and non-standardised file structures belonging to various departments dumped together that need thorough cleaning (P26.11) While performing classification with machines, it is necessary to have data points with good representation in each category (P26.12) Ongoing monitoring is a part of the rubric that includes tests for model "staleness," which can occur when the distribution of incoming data changes over time. One of the products we tested included the facility to weight training data according to its age		
G34	(P23.7) Assuming that the data sets used in AI are collected ethically to begin with, three features (training set, training set, and potential issues in matching the two) need to be carefully audited for inaccuracies, biases (P23.8) Run the data sets against a checklist of possible foibles before deployment. Is the entire space of possible data points defined and is there a reasonable presence (or understood absence) of points in some corners (such as an edge subset representing a	Data audit	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
	minority)?		
G35	(P5.5) Producing fair models (algorithms) require open and transparent definition of the populations to be protected (P14.3) Risk assessments based on relating a case to cases with similar characteristics and deducing similar outcomes is based on probabilities and the legal basis of deciding based on a likelihood is problematic (P17.3) We are not influenced by emotions. Decisions are made based on laws, rules and regulations (P17.6) Contact with citizens is made via a process governed by rules and procedures to safeguard fair and uniform decision making (P17.8) when decisions are standardized, rules and procedures are followed, and potential biases and preferences are eliminated (P24.8) To establish trust in decision-making when a person is subject to a negative decision outcome, it is crucial that the procedure of a decision is fair. Perceived fairness is relevant for trust when the decision outcome matters more to the recipient (P25.5) Perhaps the most ambitious use of AI would be to tackle issues of equality and fairness in governmental systems in a profound and transformative way	Fairness	G35; G36: **Fairness and Non-discrimination**
G36	(P12.9) People are most concerned about bias in the application of algorithms and direct or indirect discrimination is considered as one of the most crucial challenges in the use of AI-driven tools for decision-making areas	Direct/indirect discrimination	
G37	(P9.6b)) the online portal was criticized for being complex and hard to use (P9.9a) Citizens' distress aggravated by system's poor interface	User friendliness	

Group	First-order concepts (First-order Codes)	Second-order Themes	Aggregate Dimensions
G38 G39	(P10.15) Ensuring state-owned AI makes accurate predictions/recommendations is important especially in high-stakes contexts (P14.2) Technical accuracy is also a challenge (e.g. generating false negatives, false positives). Accuracy is key in high-stakes decisions	Technical accuracy	G37: G38: G39; G40; G41: **Robustness**
	(P13.3) Trustworthiness of algorithms depends on their reliability. / Public organizations must demonstrate the reliability of the algorithms they use	Reliability	
G40	(P20.2c) To plan training (and retraining) appropriately, data scientists and case workers regularly communicate regarding analyzing models' performance and new kinds of incoming data. This process supports employees' mutual understanding of how the models arrive at specific results	Model performance evaluation	
G41	(P1.5) From monopolistic algorithms and datasets to competing algorithms and datasets (P25.7) Develop AI models that incorporate different sources of data and combine insights from a range of models (so-called ensemble learning) aimed at the needs of different societal groups to curb biases and inequalities	Ensemble learning	
G42	(P7.1) Risk assessment tools (like COMPAS) must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations (P11.7) Concerns were raised that a user's conditions and doctor's diagnosis are not enough to recommend appropriate care plans and that other factors such as economic conditions, etc. must be taken into account	Context-sensitive algorithmization	G42: **Context sensitivity**