Communications on Fairness Assessments in the Public Sector

An Analysis of Roles, Skills, and Responsibilities from Dutch Use-Cases

ANONYMOUS AUTHOR(S)

Algorithms used in the public sector, e.g., for allocating social benefits or predicting fraud, often require involvement from multiple stakeholders at various phases of the algorithm's life-cycle. This paper focuses on the communication issues between diverse stakeholders that can hinder fairness assessment and potentially lead to misinterpretation and misuse of algorithmic systems. With this research objective, we conducted ethnographic research via 11 semi-structured in-depth interviews with practitioners working on algorithmic systems in the Dutch public sector, at local and national level. Applying qualitative coding analysis, we identify key elements of the communication processes that underlie fairness-related human decisions. More specifically, we analyze the division of roles and tasks, the required skills, and the challenges perceived by diverse stakeholders. Three general patterns emerge from the coding analysis. First, policymakers, civil servants, and domain experts are less involved compared to developers throughout the different phases of a system's life-cycle. This leads to developers taking on the role of advisor and decision-maker, while they potentially miss the required skills. Second, end-users often lack the technical skills to interpret a system's output, and rely on actors taking on developer roles for making decisions concerning fairness issues. Third, citizens are structurally absent throughout the algorithm's life-cycle. This may lead to unbalanced fairness assessments that do not include key input from relevant stakeholders. We formalize the underlying communication processes in a conceptual framework — introducing the *phase-actor-role-task-skill* (PARTS) ontology — that can both (i) represent the communication patterns identified in the interviews, and (ii) explicitly outline missing elements in communication patterns exhibited by actors having different roles, skills, and responsibilities in a certain use case. The framework can be further extended to explore communication issues in other use cases, design potential solutions, and organize accountability with a common vocabulary.

CCS Concepts: • **Software and its engineering** → **Risk management**; *Collaboration in software development*; • **Social and professional topics** → *Computing / technology policy*.

Additional Key Words and Phrases: Communication Framework, Fairness, Transparency, Accountability, Public Sector, Qualitative User Study

ACM Reference Format:

Anonymous Author(s). 2023. Communications on Fairness Assessments in the Public Sector: An Analysis of Roles, Skills, and Responsibilities from Dutch Use-Cases. In *Boston 2023: ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, October 30 – November 1, Boston, MA, USA.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Algorithms are increasingly being used for predicting various forms of public sector services such as allocating benefits in the domains of education, (mental) health, and detecting fraud in allowances and taxes [26, 37, 38, 51, 53]. These applications can be beneficial, but can also have detrimental consequences for citizens when used in high-stake scenarios, such as fraud detection and risk assessment. Notorious cases where incorrect predictions led to wrongful accusations of citizen minorities are the COMPAS case in the US [1], the SyRI-case [2] and the Childcare Benefit Scandal in the Netherlands[3], which resulted in lawsuits and the latter eventually led to the resignation of the Dutch government[4]. Nowadays, the problem of *fairness* in AI is widely recognized in well-established legal and ethical guidelines [10, 11, 15]. Fairness, in this context, typically refers to "fair outcomes", a principle which indicates an absence of prejudice or favoritism toward an individual or group based on their inherent or acquired characteristics through algorithmic decision-making [34]. Most of the proposed guidelines remain generic, and the tools and methods to apply them in real-world applications are missing [1, 18]. Especially, in the public sector, a disconnection exists between the proposed frameworks and the additional ethical and legal frameworks that public practitioners already deal with [17]. Furthermore, data literacy at public organizations may also not be mature enough to fully recognize ethical issues in data practices [43].

Contemporary research on fairness in AI has predominantly taken a technical approach, addressing model (dis)-functionalities without taking into account the wider network of stakeholders [7, 44]. However, algorithms are always the result of a process driven by (human) stakeholders' choices and norms [48]. A solely technical approach is insufficient in addressing all algorithm design choices, especially when they have an impact and involve codifying crucial social values, requiring to take normative decisions [33]. Decisions about which fairness criteria are assessed and how can be made throughout the algorithms' life-cycle, from data-oriented phases (eg. data collection and labeling), to model-oriented phases (eg. feature engineering, training), and execution-oriented phases (eg. deployment) [2, 25]. For example, when allocating social benefits, it has to be decided which data features represent 'eligibility' for a social benefit [4]. Also, the punitive (detecting a crime) or assistive (allocating a benefit) nature of the intervention, might require choosing different fairness metrics [38, 41]. Involving diverse stakeholders is also a design choice.[5] Particularly in the public sector, the need for involving diverse actors and stakeholders in algorithm development is important for ensuring that public interest is prioritized, and potential harms are minimized [46].

In the context of algorithmic decision-making, fairness assessment is therefore understood as a process of evaluating the extent to which algorithms and their outputs are free from bias, discrimination, and other forms of unfairness towards different groups, individuals, or communities [4, 12, 20]. In this paper, we take a socio-technical perspective on the algorithmic system and focus on the choices and practices applied at each phase of the algorithm's life-cycle that can lead to undesirable effects of bias and unfair outcomes [48]. We will analyze the internal communication processes[6] involved in fairness-related human decisions, by identifying the roles, the divisions of tasks, the required skills, and the potential communication challenges between diverse actors occurring throughout the algorithm's life cycle.

The research questions we address are the following:

- *RQ1. Which actors, roles, tasks, and phases can be identified between multi-stakeholder interactions throughout the algorithm's life-cycle?*
- *RQ2. How is information about fairness assessments communicated between diverse actors involved throughout the algorithm's life-cycle?*

[1]Correctional Offender Management Profiling for Alternative Sanctions (COMPAS): the software used to predict the risk of a person recommitting a crime was more inclined to falsely accuse African-American offenders than Caucasian offenders, see for instance [16, 34]

[2]SyRI legislation in breach of European Convention on Human Rights. https://edu.nl/xjubf

[3]The Dutch childcare benefit scandal, institutional racism, and algorithms. European parliamentary questions https://edu.nl/y3h3j.

[4]Dutch Government resigns over Child Benefit Scandal https://www.theguardian.com/world/2021/jan/15/dutch-government-resigns-over-child-benefits-scandal

[5]According to the European Commission's Ethics guidelines on trustworthy AI [15], an important step in supporting AI fairness assessment includes involving and educating all stakeholders about their roles and needs throughout the AI system's life-cycle.

[6]With internal communication we refer to communication between collaborating partners such as decision-makers, end users, developers, etc. as opposed to external communication which refers to communication with the general public [17].

- *RQ3. Which communication challenges can be identified for fairness assessment processes?*

To answer these questions, we conducted 11 interviews with practitioners working on algorithmic systems in the Dutch public sector for fraud detection and risk assessment. We identify who makes decisions about what, and in which phase of the algorithm's life-cycle. Additionally, we also look for the (co-)occurrences of mentioned actors, roles, tasks, phases, and challenges and labeled them through *in-vivo*, descriptive, and process qualitative coding analysis [40]. From the emerging coding patterns, we identified a lack of clarity in communication among different roles and actors, a lack of feedback regarding the interpretation of the model outcome, and a lack of involvement of the relevant roles and actors at the right phase. These communication issues indicate inadequate model governance such as missing roles, skills, and information exchange which can lead to misinterpretation of model outcomes, and potentially to the inability to recognize and address fairness issues throughout the algorithm's lifecycle. Next to this descriptive effort, we build the *phase-actor-role-task-skill* (PARTS) conceptual framework to characterize the structure of the communication processes based on the coding results. The framework is provided in the form of an ontology to structure the relations between the emerging coding patterns. PARTS enables to specify the relations and interactions between the identified concepts. Moreover, PARTS is introduced with the longer-term objective to be reused and possibly extended with other use cases concerning multi-stakeholder collaborations, and to promote a common vocabulary around fairness assessments.

2 RELATED WORK

Frameworks and theories from various domains have been proposed to characterize the dynamics of interactions amongst a network of actors [32, 39]. Actor-Network-Theory (ANT) and mediation theory, for example, describe the relations and interactions within a network of (artificial and natural) actors [30–32]. Following ANT, interaction with technology is never neutral as it influences or mediates the way we carry out our tasks. On the other hand, technology is continuously mediated by our social aspects, e.g. in formulating goals and other design decisions. Another related approach to describe the context of reciprocal interactions between human actors and technology is that of the socio-technical systems (STS) [39]. STS is not reserved for technology alone, it rather stresses the interactive nature of social and technical structures within an organization and society as a whole. This term is increasingly used in the field of AI to assess fairness and ethics from a broader normative context in which actors interact and operate as opposed to focusing on individual actors alone [8, 14].

Additionally, frameworks have been proposed to investigate networks and power structures. Following the tripartite model for ethics in technology, three main roles can often be identified through their responsibilities: the *developer*, who goes about the technical input, the *user* who goes about the use of the system, and the *regulator*'s role, responsible of taking the "value" decisions [36]. Research on the use of automated systems for public decision-making has shown to shift discretionary power from the regulator roles to system analysts and software designers, often making them the main decision maker [6]. Whereas decision-makers from public organizations are often involved in the procurement and deployment phase, developers, sometimes from third parties, tend to be more involved in the development phase [52]. When developers become the main decision-makers for design decisions then this can exclude those stakeholders without technical knowledge in important value decisions for the system [13, 29]. These imbalanced power dynamics can ultimately lead to a form of technocracy, where governance and (moral) decision-making are based on technological insights and may only yield technological "solutions"[36].

In the Dutch public sector, multiple stakeholders from private and public organizations often collaborate at various phases of the algorithm's life-cycle [26, 45]. Stakeholders might be involved at different moments and places depending on their roles and tasks and therefore responsibilities can become diffused across a network of multiple actors simultaneously [5, 52]. Related empirical field research has been conducted to investigate data practices of Dutch local governments [17, 28, 43]. For example, Siffels et al. (2022) argue that with the process of decentralization, many tasks from the central government were delegated to municipalities without giving them more resources and capacities. As a result, municipalities use data practices to deal with additional tasks and to distribute limited (social) resources. Due to a lack of data literacy, public servants are unable to recognize ethical issues and thus seek collaboration with external partners which affects their ability to be transparent and responsible regarding their data projects [43]. The lack of transparency and responsibility in outsourcing parts of the algorithms' life-cycle is often also referred to as the "problem of many hands" [9]. In another field research, Jonk and Iren (2021) performed semi-structured interviews with key personnel and decision-makers at 8 Dutch municipal organizations to investigate the actual and intended use of algorithms [28]. They found that there is a lack of common terminology and algorithmic expertise not only at a technical but also at a governance and operational level. The authors argue that municipalities would benefit from a governance framework to guide them in the use of tools, methods, and good practices to handle potential risks. Furthermore, Fest, Wieringa, and Wagner (2022) investigated how higher-level ethical and legal frameworks influence everyday practices for data and algorithms used in the Dutch public sector [17]. They investigated public sector data professionals at Dutch municipalities and the Netherlands Police. They found that the practicality of proposed frameworks remains a challenge for practitioners because they typically do not feel competent or miss the required skill set to make decisions regarding responsible and accountable data practices. Data professionals, as a result, get too much autonomy and discretion in handling questions that belong at the core of public sector operations. The authors argue that efforts need to be put into the operation and systematization of legal and ethical questions across the data science project life-cycle. Research on public algorithms has also been conducted in the fields of Human-Computer Interaction (HCI), Science and Technology Studies (STS), and Public Administration (PA) [27, 33, 42]. For instance, Saxena et al. developed a framework for high-stakes algorithmic decision-making in the public sector (ADMAPS) where they qualitatively coded data from in-depth ethnographic study on the daily practices of U.S. Child-welfare caseworkers and prior literature [42].

Lastly, tools such as ontologies [7] can be used to characterize communication structures and knowledge exchange between actors[24]. Ontologies are common tools to structure knowledge and to promote a common vocabulary within a certain domain [35]. Earlier ontologies on AI risks and fairness have been proposed [19, 23]. AIRO for example is an ontology that underpins the AI Act, a regulation proposed by the European Commission to tackle various forms of negative impact caused by the misuse of AI [10, 23]. The Fairness metrics ontology (FMO) structures knowledge for fairness notions, metrics, and the relations between them [19]. We propose that what is still missing in both ontologies is a clear framework to characterize the communication (issues) between diverse actors with different skill sets that underlie fairness-related human decisions.

We propose that what is still missing in previous work is a clear framework to characterize the communication (issues) between diverse actors with different skill sets that underlie fairness-related human decisions throughout the algorithm's life-cycle. Following the above frameworks and theories, we assume a perspective of a socio-technical interactive network, in which fairness assessments for algorithms are understood as part of a governance structure where actors with different roles interact. We conduct interviews to specifically focus on interactions by means of internal communication exchange and challenges that might arise between diverse actors in the Dutch public sector.

[7]An ontology can be understood from a philosophical perspective referring to "the nature and structure of reality". From a knowledge engineering perspective, however, it refers to the modeling of a structure of a system — by organizing relevant concepts and relations [24]. They are also described as "conceptual schemas", "formal specifications of a conceptualization" and as "the abstract and simplified view of the world we wish to represent and describe in a language that is understandable by humans and/or by software agents" [3, 24].

The communication challenges are finally characterized in a conceptual framework in the form of an ontology. With this conceptual framework, we aim to provide a structured approach to the planning and auditing of communication processes around technical-decision making in the public sector.

3 METHODOLOGY

3.1 Semi-structured interview

We conducted 11 semi-structured interviews, with each one divided into three sections and lasting an hour. We formulated the qualitative interview questions in an open-ended manner, where participants were able to share their information in their own words whilst following a general structure of topics [21, 22]. The questions are added for reference in table 4 in the appendix. The questions can be divided into three main categories:

1 **General**: questions about the topic of the use-case, actors involved, and the content of the respondent's (teams) role. The goal of the system is identified, as well as the envisioned (end) users.
2 **Development process**: type of input, resources, tasks, and roles needed throughout the development process to make informed decisions. The phases of the algorithm's lifecycle are investigated by mapping out the tasks made in 5 phases: formulation, development, evaluation (go-no-go), deployment, and monitoring phase.
3 **Considerations**: perceived challenges for role and task division, potential improvements or failures of the system, and communication gaps are identified. Questions about assessing errors, bias, as well as the potential negative impact of the model, are asked explicitly.

We start by giving interviewees the opportunity to mention internal communication and fairness issues spontaneously and on their own terms. The first two sections ask interviewees to describe rather generic procedures and practices in the system's life-cycle. We then precisely ask about communication issues in the third part of the interview. We preliminary tested the interview questions with a pilot involving colleague researchers from different disciplines. The questions were deemed suitable for letting interviewees describe their process of communication and related issues. The suitability of the questions was checked in terms of comprehensibility and relevance to the research questions. No questions were altered afterward.

3.2 Case studies

We recruited participants who have been collaborating in multi-stakeholder projects in the Dutch public sector. We specifically targeted participants working in the domains of fraud detection and risk assessment, using a repository of use cases[8] from the Dutch Ministry of interior affairs[9] and the snowball sampling technique. These use cases were of particular interest because the outcomes can directly affect citizens.[10].

For this research, *(N = 11)* interviews were conducted with (*N=10*) actors involved in Dutch social domain use cases, and *(N = 1)* actor in the Dutch educational domain. Table 1 describes the participants (PP), the use case, the role they filled at the time of involvement and if they have technical skills. With technical skills we refer to those that are (not) educated or have (no) experience in technical science. Topics discussed for the use cases were predicting fraud amongst citizens (eg. when applying for public service) *(N = 5)* or predicting the need for social benefits amongst citizens in

[8]Some examples of public domain use cases in the Netherlands can also be found via the Artificial Intelligence Netherlands Coalition (NL AI Coalitie) website https://nlaic.com/use-cases and in [26, 51].
[9]Dutch Ministry of Interior Affairs and Kingdom Relations https://www.rijksoverheid.nl/ministeries/ministerie-van-binnenlandse-zaken-en-koninkrijksrelaties
[10]To further confirm use-case relevance, in 2019 inspection and enforcement was identified as the biggest category for AI usage in the Dutch social domain which refers to the prediction of (security) risks by identifying patterns of behavior [51].

Table 1. Interviewees description

Interviewee	Role	Technical background	Use case domain
PP1	Developer & Researcher	yes	Risk assessment
PP2	Program Manager	no	Risk assessment
PP3	Program Manager	no	Risk assessment
PP4	Product Owner	yes	Fraud detection
PP5	Advisor & Social researcher	no	Risk assessment
PP6	Researcher & Developer	yes	Education: face expression detection
PP7	CTO development team	no	Risk assessment
PP8	Advisor & Data science researcher	yes	Fraud detection
PP9	Developer & Researcher	yes	Fraud detection
PP10	Advisor & Researcher	no	Fraud detection
PP11	Innovation manager	yes	Fraud detection

the coming years *(N = 5)*. Albeit the difference between risk and fraud cases is not always clear-cut (a fraud detection algorithm can also be used for risk and vice-versa). The categorization of the use cases is therefore based on the punitive (fraud detection) or assistive nature (risk assessment) nature of the policy measure. Before conducting the interviews, participants received some example questions and a short description of the research. At the start of the interview, participants gave their consent for collaboration.

3.3 Qualitative coding analysis

To provide insights into the role and task division amongst actors in multi-stakeholder projects, we performed qualitative coding analysis to label key *codes* from the interview output [11]. We use in vivo [12], descriptive [13] and process coding [14] to identify the process of communication exchange between diverse actors, as well as the practices and choices made at each stage of the algorithm's life-cycle. Beforehand, both analysts agreed that particular attention should be drawn to describing the role of the stakeholders, depending on the task, phase, and on potential challenges that might occur. For example, if a participant were to mention that "person X is a developer and performs bias analysis in the development phase" we label the actor, the role, the task, and the phase. We also pay attention to the direction of information exchange between actors (eg. person X hands over model outcomes to person Y). Furthermore, communication challenges and information that was repeated or pointed out by participants as essential for fairness assessment were coded. For the sake of supporting a common terminology, we took a broad range for the formulation of the codes. For example, participants mentions for engineers, coders, and data scientists are all considered as the role of developers. Also, for

[11] Saldaña (2013) describes "A code in qualitative inquiry is most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data." In addition, a code can be understood as a researcher-generated construct that symbolizes the construct and assigns an interpreted meaning .

[12] In Vivo coding is also named "literal coding" and refers to a word or short phrase from the actual language found in the qualitative data record eg. terms used by participants themselves [40, 47].

[13] descriptive coding refers to summarizing the basic topic of a passage of qualitative coding in a word (noun) or short phrase [40].

[14] process coding refers to "action coding" which implies action from more simple observable activity (eg. reading) to more general conceptual action (such as adapting) [40].

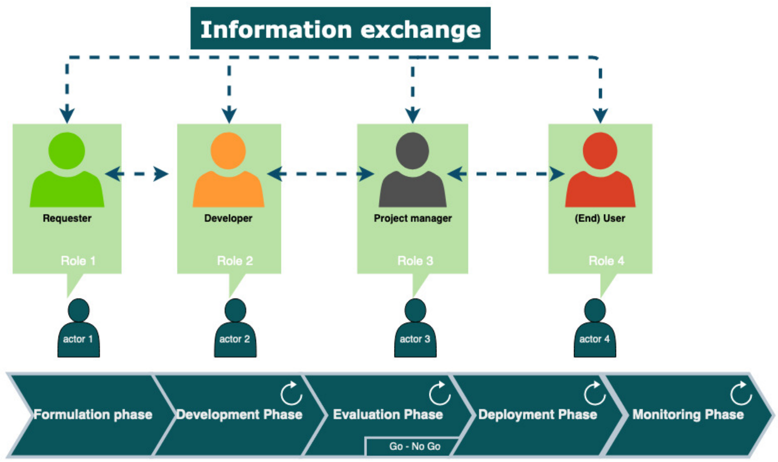

Fig. 1. Key phases of a system's life-cycle, and their *actors* and *roles*, identified from the interview transcripts.

the phases, we considered the testing phase, modeling, and experimentation phase as the development phase. These are grouped because they were not purposely separated but used interchangeably by participants in the interviews. In figure 1, an overview can be found of examples for the identified key phases, actors and roles from the interview transcripts.

The coding analysis is done in multiple cycles, where in each round of coding pieces of text are merged or split into categories. Two of the authors perform a separate coding analysis, to reduce the impact of personal bias. One performed the coding analysis by hand whilst the other used a coding analysis tool [15]. Each analyst then identified the roles, tasks, phases, and challenges from the interview transcript independently. We finally compared both coding analyses to identify discrepancies or alignments, as described in the Results section. Mentions of coding are done by counting the (co-)occurrence for roles and phases, roles and tasks, and roles and challenges.

3.4 Constructing a Conceptual framework

Based on the codes identified with the qualitative coding analysis from the interview transcripts, we organize the main elements in a conceptual framework for the general communication processes between diverse actors that underlie fairness-related decisions. We build the conceptual framework using the method of constructing an ontology. The framework is therefore built in an iterative manner presuming that it can be revised and extended for future work when more input is available. We follow the work by Noy & McGuinness (2001)[16] on creating ontologies [35] and take other ontologies and conceptual frameworks as a reference [19, 23, 50]. The ontology promotes a common terminology so that diverse actors can contribute to creating transparency and accountability throughout their system's development practices.

A **Creating concepts from codes** We use the main identified codes from the qualitative coding analysis to make concepts, also known as classes, in our ontology. As mentioned in the previous sections, we focus on identifying roles, actors, tasks, phases, and challenges. We purposely separate actors from roles to see how many roles

[15]Atlas.ti: The Qualitative Data Analysis & Research Software https://atlasti.com/
[16]"Ontology development 101: A Guide to Creating Your First Ontology"

actors can take on for a certain task. These concepts are the main focus of the ontology for representing the communication structure.

B **Defining Descriptions** We add descriptions to each concept to agree on common definitions. Descriptions are based on answers we got from the interview transcripts, the definitions we found from documents provided by the European Commission on Trustworthy AI, and from other sources in the literature [11, 15, 23, 36, 49, 50].

C **Defining properties and relations between concepts** We describe the internal structure of the concepts by adding properties. For instance, for actor, we use a property to describe their public or private affiliation. We determine the relations between the concepts to describe the generic process of communication. For example, an actor always takes on a certain role and is involved during a certain phase, when performing a certain task. Since we identified various challenges from the interview transcripts, we decide that challenges can occur for all the other defined concepts.

D **Competency Questions** We determine the scope of the ontology by thinking of competency questions that a knowledge base derived from the ontology should be able to answer. We use our research questions as guidance to answer and formulate competency questions for who (actors, roles) does what (tasks) when (phase), and which information exchange is needed for that.

E **Creating instances**: We create generic instances of concepts for the main concepts. By doing this step, we aim to describe the concepts in relation to each other. This is done by taking a phrase from the interview transcript eg. a phrase that contains a challenge, and filling in concepts for specific slot values. We go back to the other steps if the instance calls for revision of the ontology.

4 RESULTS

4.1 Semi-structured interview

For the majority of use cases *(N = 10)*, the procurement for the algorithm came from government organizations and municipalities. Furthermore, in all use cases, there were multiple stakeholders involved with varying technical expertise—from social workers to developers, researchers, program managers, and third parties. Finally, in most use cases *(N = 10)*, the envisioned end-users of the systems were policymakers or other domain experts at municipalities with minimal or no technical expertise. There was one exception for the educational domain, which envisioned teachers as end-users. 6 interviewees mentioned having a technical background themselves, but not necessarily in AI or software development.

4.2 Qualitative coding analysis

4.2.1 Qualifying the stakeholders. Figures 2, 3, and 4 illustrate which roles were mentioned the most based on co-occurrence per Phase, Task, and Challenge. Overall, the developer role was mentioned the most *(N = 189)*, followed by End-users *(N = 107)* and Policy Makers *(N= 92)*. More formulation of roles, phases, tasks, and challenges were identified but mentioned less and thus not shown here. Figure 2 shows that developers are most prominent in the development, evaluation, and formulation phase and less in the deployment and monitoring phases. PP1, who filled a developer role reported *"we don't monitor what the municipalities are doing with the results."* and *" feedback is needed on how the results will be used in deployment"*. End-user and policy-makers are reportedly the same in almost all use cases. They are also the second highest in occurrences for phases. Moreover, figure 2 demonstrates that the monitoring phase *N = 10* is mentioned the least throughout the interviews whereas the evaluation phase is mentioned the most *N = 90.*

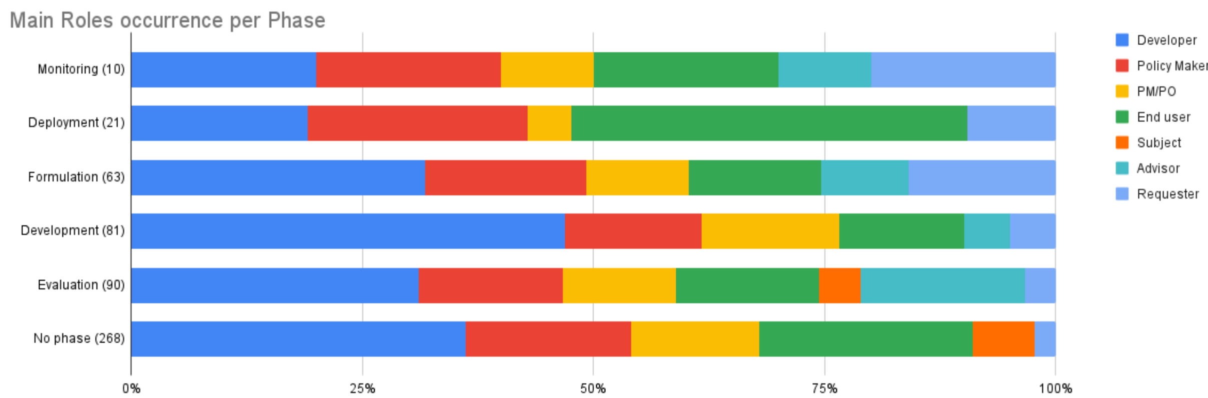

Fig. 2. Mentions of Roles (relative) per Phases. The total amount of occurrences per phase is mentioned on the y-axis (from least to most). Note that the developer role (blue) is mentioned most in all phases, followed by end-users(green) and policy-makers(red) and subjects (orange) least.

The subject role occurs the least for phase and task and is in almost all cases filled by citizens. Participants reported the need for more citizen involvement and being more transparent to citizens throughout the phases algorithms' life-cycle. PP2, a program manager, mentioned that *"the plan is to check eg. with research labs how citizens can give feedback on the model. But currently, nothing is envisaged"*. On involving citizens, another researcher (PP10) working on fraud detection, reported that *"it depends on the type of AI. If it has an impact on citizens or uses a lot of data from citizens, it would be relevant to include a focus group of citizens from the beginning but it is less relevant for road repairs."* Other roles are also mentioned that need to be more involved throughout the phases. PP9, a developer, reported that *"For the future, we could incorporate stakeholders at earlier stages in the development to see what the potential sources of bias are."*

Additionally, requester roles are not mentioned frequently and are mostly involved in the formulation phase for funding or initiating the project. A requester often resides at external governmental organizations such as ministries and collaborates with municipalities. The subject role as well as the requester role were never described as end-users.

In figure 3 it is shown that the developer role is mentioned most for tasks, e.g. technical decision-making, researching, bias analysis, and consulting, and the least for model usage. This indicates that actors taking on developer roles are the most prominent in making decisions throughout the algorithm's life-cycle. On what type of decisions one makes, a PP9 who filled a developer role at the time, says that they *"decided on how to improve accuracy and handling issues. For instance, gathering more data diverse to handle bias."*. Developer PP1 mentioned that they *"define and chose metrics for the models."* and that these*"are defined in collaboration with the municipality but choosing metrics and trimming down after input was decided by the two of their team."*

Program managers and product owners (PM/PO) often work in the same research teams as developers and are either hired externally or internally by a public requester. They are the main decision-makers for involving (external) stakeholders such as advisors, requesters, and end-users, and sometimes being the only ones in direct contact with them to answer questions. PM/PO's are often reported to supervise developers in technical decision-making but rely on their judgment for bias, fairness, and risk factor analysis. Program manager (PP2) said that for handling error rates and biases they *"rely on the technical teams' judgment"*. Also they mentioned that *"the technical colleagues give advice when the model is good enough, but it's a bit of a grey area. We also rely on literature"*. PP3 mentioned that they *"don't go*

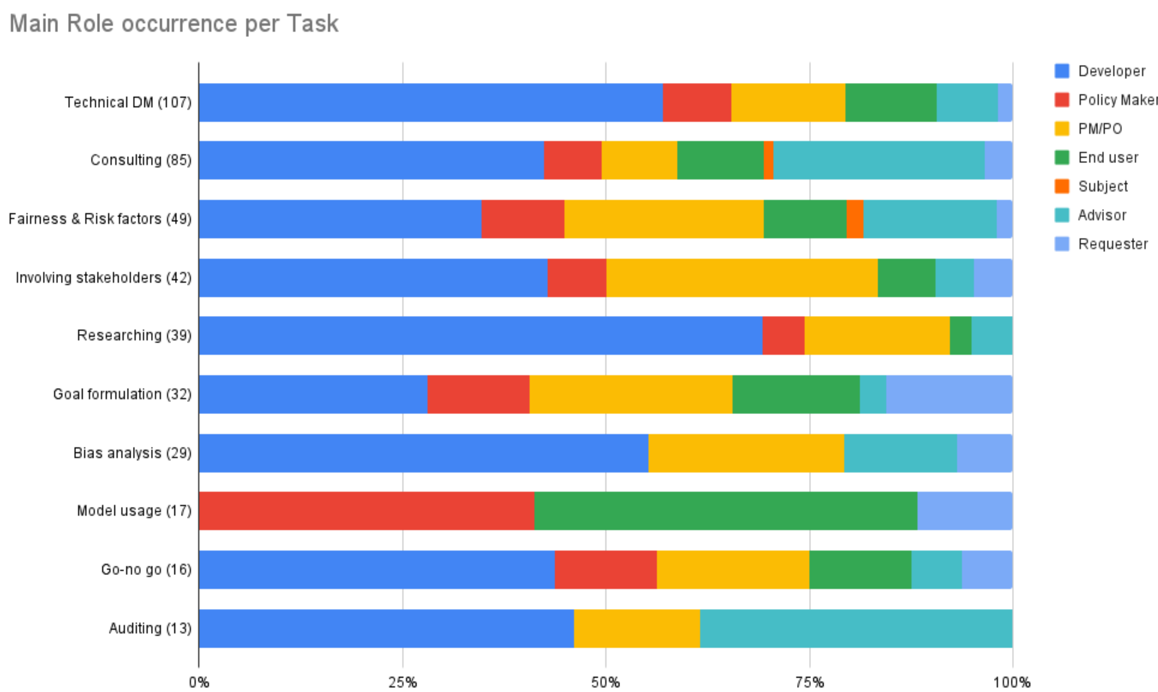

Fig. 3. Mentions of Roles (relative) per Task. The total amount of occurrences per task is mentioned on the y-axis (from most to least). Note that technical decision-making is mentioned the most. Moreover, the developer role (blue) is mentioned most for all tasks except for model usage.

to users" for error or bias analysis but *"talk to the data scientist".* Product owner (PP4) said that *"it is time intensive to explain [bias analysis] to stakeholder users. Bias analysis is sometimes so complex, even as an expert I sometimes don't understand it, and it takes a lot of time".*

Advisors are often mentioned to consult on the system in the evaluation phase before deployment, or when the project is halted. Advisors are presented as giving advice on 1) domain knowledge, 2) technical knowledge, or 3) ethical knowledge. Advisor roles are sometimes combined with developer roles. A third-party developer can be hired to analyze the code, give technical advice or even build the model. PP3 added that they *"hired an external bureau for auditing and investigating the algorithm".* Also because they *"could not get reliable predictions because the social domain changes all the time, and it's hard to keep track of these changes—for example in social support—and how that impacts the system".* PP10 mentioned that *"an external company was hired to develop the model for the municipality"*, which made the *"data ecosystem quite complex".* A social domain expert and researcher at the municipality (PP5) mentioned that they *"were involved to give feedback as an involved bystander. But it was hard for someone like me to understand what the difference between implementation and design is and what that means for real-life implications".*

In figure 4 the relative occurrences for roles per challenge are demonstrated. Most communication challenges were reported amongst end-users/policymakers and developers due to lack of interpretation, missing involvement, and (risk) oversight in a certain phase. Program manager (PP3) mentioned that is a challenge that *"we don't know if governments*

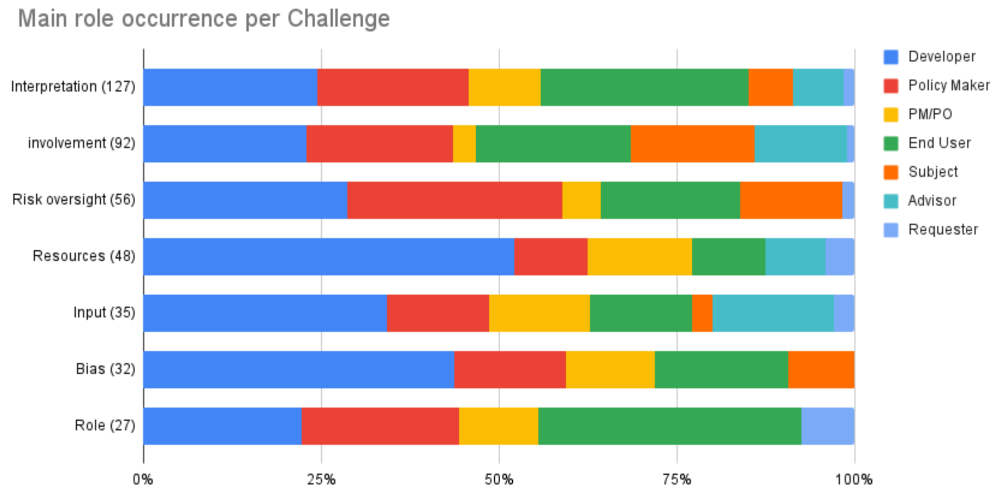

Fig. 4. Mentions of Roles (relative) per Challenge. The total amount of occurrences per challenge is mentioned on the y-axis (from most to least). Note that interpretation and involvement issues were mentioned most. Moreover, developer (blue) is mentioned most followed by end-users (green) and policy-makers (red).

and municipalities can understand the model". Also, PP1 mentioned with regard to model usage at the municipality that *"it's hard to get a focused answer on how they are going to use the model and what the results will be."* Therefore, more input is needed on the interpretation and use of the results from end-users in the deployment phase. Vice versa on feature selection and bias analysis in the development phase end-users require input from developers. On the matter of involvement, another program manager (PP2) reported that *"More frequent and streamlined collaboration with the municipality is needed"* and that they *"would like to have closer contact with the municipality."*. PP1 also confirmed that in their case *"the municipality is too loosely involved in the project, more involvement is needed"* when asking about challenges regarding role division.

End-users and policy-makers are often reported to miss the technical skills to understand the uncertainty of predictions and limitations of the model in real-world settings. For example, PP10 mentioned that the most difficult challenge is the *"gap between data scientists and policymakers. How to make sure that what is developed is being well understood and useful for those of non-tech background."*. PP1 also confirmed that the *"Most important risk is that the model will not be used or is misinterpreted. For example, mixing up correlation and causality might lead to not helping people at risk of poverty."*. PP6 mentioned that *"People could trust the model blindly and mistake it for a decision-making tool".* and lastly, PP8: *"Not sure if the inspectors fully understood why certain cases were flagged as misuses or put on the list."*

With regard to biased decisions, another participant (PP9) mentioned that *"there should be more focus on asking users what policymakers perceive as risks and biases"* and that it is *"difficult for them to understand that there are many different interpretations. What it really means to be a 'true positive', is this person really a fraud, or was this person not able to fill in the forms properly?".* Misinterpretation and misuse of the system by end-users and policy-makers were mentioned most in the monitoring and deployment phases. PP6 confirms that *"training for users is needed, to remind users not to rely on* *the tool but that the decision is up to them."*. This also explains why policy-makers and end-users are mentioned a lot throughout all phases because of their co-occurrence with challenges in certain phases.

Program managers and product owners are also hesitant to involve citizens for a lack of appropriate frameworks, and potentially hostile attitudes. "*There is a long history with the citizen council for consultation and it is usually conflict-based. It's hard to make fruitful collaboration, getting them to understand the issues and getting them out of anger mode."* said a product owner (PP4) working at the municipality. PP7, on previous involvement of a citizen council: *"they said no on the feasibility of the model from the municipality. They did not get it. It was more of a general no to technology instead of asking a targeted question".*

4.2.2 Qualifying the communication processes. Our qualitative coding analysis shows the importance of differentiating individual stakeholders by their *roles* and *skills*, meaning, having specific skills to match particular roles diverse actors need to have within the process. Fairness assessments are impacted when individual *actors* with specific *roles* or *skills* are missing in the communication framework.

For example, *actors* with technical *skills* often assume the *role* of policy-makers. Yet, they may miss the required *skills* in the domain of application. For instance, the developers of a fraud detection system may decide on error metrics, bias issues, and test sets to measure them. Yet, they may not have the expertise in demographics, finance, and forensics to ensure that their technical decisions are fair or allow them to assess fairness outcomes appropriately. On the other hand, the policy-maker role may not have the technical skills to understand bias or error metrics. In this case, we can consider that *actors* with the *role* of policy-makers are missing. Also, *actors* with practical *skills* in the domain of fraud are missing.

As we reflected on our qualitative coding results, specific patterns of issues in communication frameworks emerged. We were able to identify key elements of communication frameworks that are particularly critical to assessing fairness. We found that issues arise depending on how these elements are combined with other elements, to form communication frameworks with similar patterns and structures. These communication patterns are not easily described with qualitative coding analysis only. By solely counting the codes for (co-) occurrences, we could not capture the type of relation between codes that are important for determining the process of communication. For instance, a communication challenge could be that a role is missing in a certain phase, which would also count as a co-occurrence for a role within a phase. Their structures have variations that are not easy to identify in the narratives of interview transcripts. The communication patterns are also not easy to document in written form only. Therefore, we turned to ontology modeling for documenting the communication patterns we identified. In the next section, we provide a more structural analysis of the communication patterns using the PARTS ontology. PARTS was specifically designed for—but not restricted to—modeling the results of our qualitative coding analysis.

4.3 The PARTS ontology

The Phase-Actor-Role-Task-Skills (PARTS) ontology models the structure of communication frameworks and the basic elements for the communication issues identified in our qualitative coding analysis. It comprises rather generic concepts, such as *Phase, Actor* and *Information Exchange*, and concepts that are more specific to fairness assessment, such as *Role, Task*, and *Skill*.

The communication frameworks require at least 2 generic elements to materialize the communication process: the stakeholders who exchange information (represented with the concept *Actor*), and the act of communicating (represented with the concept *Information Exchange*). We found that the relevant context of the communications can be

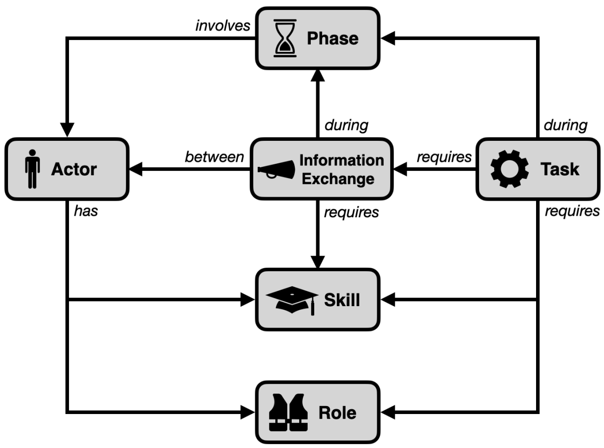

Fig. 5. Basic concepts of the PARTS ontology for characterizing the structure of communication frameworks, and the issues we identified. For example, fairness assessment *Tasks* may be missing at certain *Phases* of a system's life-cycle. Or *Actors* may not have the right *Skills* or *Role* when performing fairness assessments.

Concept	Description	Relation	Property
Task	Actions carried out by Role	Requires Role, Skill, Info exchange, during Phase	
Role	Function or a part filled by Actor	Actor Has Role, Task requires Role	
Actor	Entities that perform Tasks, actively or passively	has Role, Skill, Phase involves Actor, Info Exchange between Actor	*affiliation*: public or private, *is missing*: yes/no
Phase	indicates the evolution of the system from conception through retirement	Involves Actor, Task during Phase, Info exchange during Phase	
Skill	Professional ability, expertise, or knowledge needed in practice to successfully complete a specific task	Info exchange requires Skill, Task requires Skill, Actor has Skill	*is missing*: yes/no
Info exchange	Communication transfer between actors	requires Skill, between Actor, during Phase, Task requires Info exchange	*is missing*: yes/no

Table 2. Concept definitions

modeled with 4 additional concepts (*Phase, Task, Role, Skill*). We found that these simple concepts were sufficient for modeling generic communication challenges that we observed in several interviews. The ontology can be extended with additional concepts, e.g., to represent the affiliation of *Actors*, or the information that is exchanged. We decided that further research is needed to further refine the ontology, and that adding more concepts at this stage comes at the risk of making the ontology less applicable or generalizable to new contexts.

We thus selected these 6 concepts from the codes of our qualitative analysis of the interview transcripts. Table ?? provides the definitions and properties of these concepts, and Figure 5 shows their relationships. Figure 6 provides an example of the instances of the ontology that we derived from the interviews.

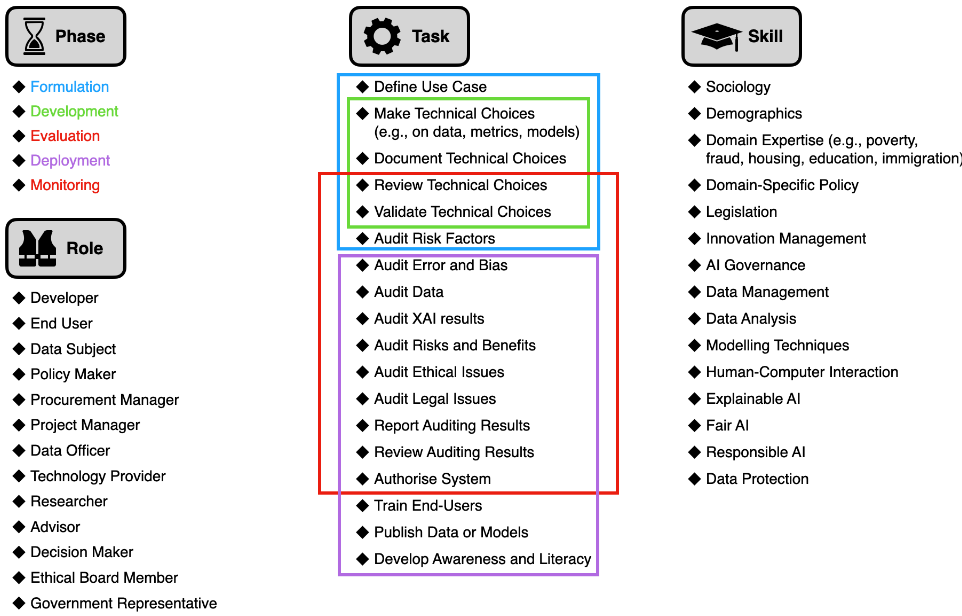

Fig. 6. Generic instances of the PARTS ontology, identified in several interviews. The generic instances of *Tasks* can occur during different *Phases*, e.g., as the colored boxes show. The *Tasks* occurring during the *Phases* of ◆ *Evaluation* and ◆ *Monitoring* can largely overlap. Instances of *Actors* are not generic, they represent private individual persons (e.g., the interviewees themselves).

Role types	Description
Developer	research, design, and/or develop algorithms
Policy-maker	responsible for designing and overseeing the carrying out of policy and social decisions
Project manager/ Product owners (PM/PO)	supervise the projects for the development of the system and oversee documentation checks and balances
End-user	(in)directly engage with the system and use algorithms within their business processes to offer products and services to others
Subject	organization or entity that is impacted by the system, service, or product
Advisor	give constructive feedback on the system throughout the life-cycle
Requestor	who are the main client and investor for the use-case

Table 3. Main roles description

We found that a few simple properties and relations between concepts were sufficient to represent the communication issues we observed in the interviews. The property *requires* assigned to the fairness assessment *Tasks* allows representation of the *Roles, Skills* and *Information Exchange* needed to perform the task. This part of the communication framework is essential to define the high-level requirements for accountability and transparency. The practical implementation of these requirements can be modeled by associating *Actors* with specific *Skills* to the *Roles* they take in fairness assessment *Tasks*.

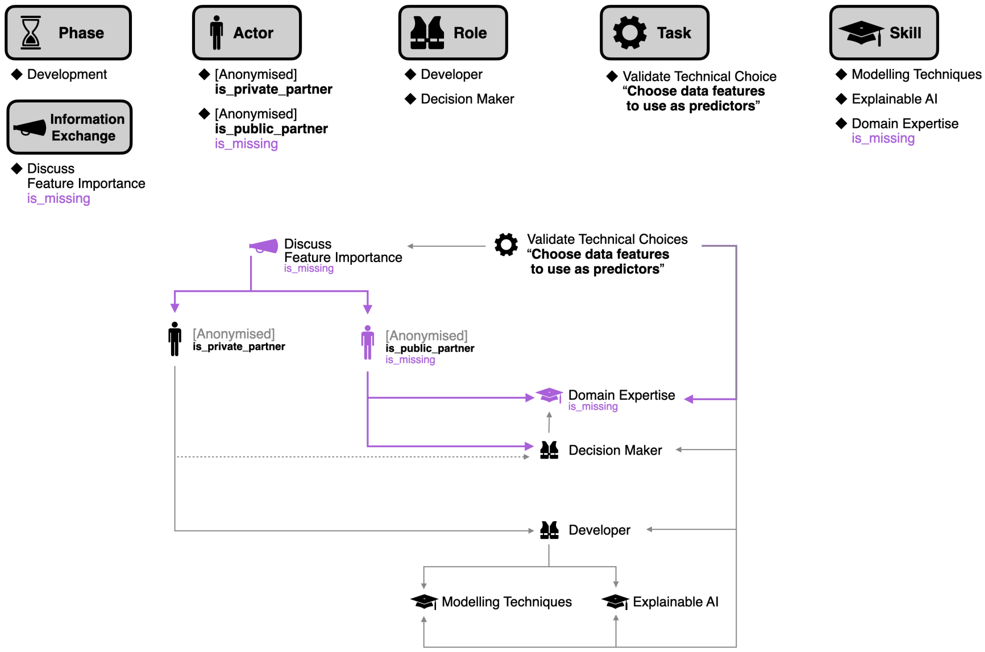

Fig. 7. Example of a generic communication challenge observed in interviews, with elements of the communication framework (top) and their structure (bottom). Existing elements are in black, and missing elements are in pink. For the *Task* of ◆ *Validating technical choices*, the *Role* of ◆ *Decision Maker* is missing an *Actor* who has *Skills* in the ◆ *Domain Expertise*, and who is independent of the ◆ *Developer*.

With the properties *during* and *involves*, the ontology can associate *Tasks, Information Exchanges* and *Actors* to the *Phases* of the system's life-cycle. This provides a generic temporal overview of the communication and assessment processes, e.g., to reflect on the assessment *Tasks* or *Actors* (e.g., stakeholders from specific institutions) that may be missing at specific *Phases*.

Our main finding is that adding the property *is_missing* to any element of the communication framework (i.e., to any of the 6 concepts) is of great interest for expressing the issues that interviewees mentioned. We represented the elements of communication frameworks using these 6 concepts precisely because issues arise if any of these types of elements are missing. For example, an *Actor* may be missing specific *Skills*, or a fairness assessment *Task* may be entirely missing. Communication issues also arise from missing a *Information Exchange*, as shown in Figure 7.

The concepts we selected, and their descriptions, are in correspondence with those provided by the European Commission on trustworthy AI, and other sources in the literature [11, 15, 23, 36, 49, 50]. For example, the developers and end-users (as instances of *Actor*) are, respectively, those who "research, design, and/or develop algorithms" and those who "(in)directly engage with the system and use algorithms within their business processes to offer products and services to others" [15]. Data subjects (also instances of *Actor*) are "an organization or entity that is impacted by the system, service or product" [23].

5 DISCUSSION

The patterns identified from the coding analysis indicate a heavy reliance on the (technical) skills of the developer role in the development, formulation, and evaluation phases. This means that developers likely take on most tasks and decisions throughout the life-cycle which may potentially influence fairness outcomes.

These results are based on the co-occurrence of mentions provided by participants in interview transcripts. However, participants tend to forget involved internal and external stakeholders in the development process or don't see some as influential for choices, practices, and protocols in the development process. This does not mean that forgetting a certain role or actor fully reflects the actual governance structure or experienced communication challenges. Participants may not have oversight, may not be willing to provide specific details, or were perhaps steered by the interview questions.

Besides, we recognize that in practice, formulations for roles and groups can vary and can be diffused. For instance, in some cases, actors identified as developers would primarily identify themselves as a researcher — who also carry out developer tasks or vice-versa. Moreover, counting (co-)occurrences is not enough to assess the nature and structure of interactions within a network. As mentioned in the results, sometimes an occurrence would be counted for role and phase when actually the content of the code would say that 'role X was missing in phase Y' and thus was more an occurrence for the challenge than for phase. Therefore, the PARTS ontology was needed to provide an overview and to characterize the interaction between roles and the challenges within a bigger network.

The patterns additionally indicate a need for increased input of end-users, policy-makers, and developers at the right phase, in particular during design choices and output interpretation. Even when there is involvement, participants indicate that the necessary skills are missing to provide for informed decisions. There is uncertainty about whether the algorithm actually provides what it is supposed to perform, coupled with weak support and guidance for users, limited supervision, advice, and a lack of feedback loop for AI developers. These findings demonstrate a growing information gap between technical experts on one side and domain experts and policy-makers on the other side. Presenting algorithms and features in an understandable way to end-users and policy-makers is increasingly needed, aimed specifically at a clear identification of potential risks, failures, and errors of the AI model.

Moreover, within patterns, there is no clear governance or process structure regarding roles, actors, information flows, and responsibilities during the algorithm's life-cycle. The findings revealed it was not always clear who makes final (mostly policy) decisions on the further development or use of algorithms, or what is the (legal, procedural, information) basis for such decisions. This seems to be due to the lack of actors filling the right roles at the right phase; which leads to actors taking on multiple roles at once they are not fully equipped for, consequently leading to a discretionary imbalance. For example, developers tend to take on multiple roles such as advisor, researcher, and decision-maker, without being fully aware of the business logic and domain expertise. In addition, policy-maker or social worker roles tend to coincide with end-user roles without being allocated a clear mandate for making informed value decisions. These findings are not new and confirm previous work done on the growing discretionary power of developers and unclear role divisions for automated decision tool usage in policy-making settings eg. [6, 36].

The issues of information gaps, unclear role divisions, and co-occurrence were characterized by the PARTS ontology. PARTS is a communication tool that can explicitly structure and outline missing elements such as missing roles, skills, and involvement. Furthermore, it promotes a common terminology beyond solely the technical perspective of fairness assessments but also on the matter of accountability. We purposefully kept the ontology as simple as possible because every public sector use case will have its own (normative) context, specific governance, and communication structure.

Also, we do not claim that fairness issues are resolved with PARTS. We solely emphasize that governance structures, roles, and communication processes between diverse actors are of importance when addressing fairness assessments.

This paper shows that when responsibilities and roles are unclear, and there is a lack of communication or miscommunication, this can lead to misinterpretation of the system output and even to misuse of the algorithmic system. Finally, the coding analysis, as well as the generic patterns presented in PARTS are based on *(N = 11)* interviews. More data is needed in the future to verify and specify these findings. Future work would require increased development and use of practical frameworks for communication, role and responsibility distribution, and governance processes.

6 CONCLUSION

This paper examines the roles of different stakeholders engaged in the process that leads to algorithm procurement and development in the Dutch public sector. In particular, we focus on the analysis of communication patterns exhibited in fairness assessments. The field research we conducted points to inadequate or missing explicit guidance or rules regarding roles, skills, and information exchange throughout the algorithm's life-cycle which leads to misinterpretation of model outcomes and potentially to unfair outcomes. The results have been characterized using the Phase-Actor-Role-Task-Skill ontology and can be used to indicate a lack of involvement and feedback between developer, end-user, and policy-maker roles. End-users and policy-makers often lack the technical skills to interpret the system's output or to estimate potential fairness issues. Therefore, they rely on the technical skills of developers for making apparent technical decisions such as bias and fairness assessments, potentially influencing policy outcomes. As a result, developers take on extra roles such as advisor and decision-maker while potentially lacking the business logic and domain expertise required. This also results in developers being the most prominent in most tasks and phases of the algorithm's lifecycle. Lastly, we observe that citizens are structurally absent throughout the algorithmic life-cycle, even though it is mentioned that their involvement is needed in the future for balanced fairness assessment. The PARTS ontology is built to describe these patterns and to be used and extended by domain experts: to audit and design frameworks to deal with communication issues, to support transparency and accountability amongst diverse actors, and to promote the creation of a common vocabulary on fairness assessments.

REFERENCES

[1] Kasun Amarasinghe, Kit T. Rodolfa, Hemank Lamba, and Rayid Ghani. 2020. Explainable Machine Learning for Public Policy: Use Cases, Gaps, and Research Directions. *CoRR* abs/2010.14374 (2020). arXiv:2010.14374 https://arxiv.org/abs/2010.14374

[2] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. 2019. Software engineering for machine learning: A case study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 291–300.

[3] Grigoris Antoniou, Enrico Franconi, and Frank Van Harmelen. 2005. Introduction to semantic web ontology languages. *Reasoning web* 3564 (2005), 1–21.

[4] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2019. *Fairness and Machine Learning: Limitations and Opportunities*. fairmlbook.org. http://www.fairmlbook.org.

[5] Mark Bovens. 2007. 182 Public Accountability. In *The Oxford Handbook of Public Management*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199226443.003.0009 arXiv:https://academic.oup.com/book/0/chapter/292909174/chapter-ag-pdf/44514193/book_34524_section_292909174.ag.pdf

[6] Mark Bovens and Stavros Zouridis. 2002. From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control. *Public Administration Review* 62, 2 (2002), 174–184. https://doi.org/10.1111/0033-3352.00168 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/0033-3352.00168

[7] Shea Brown, Jovana Davidovic, and Ali Hasan. 2021. The algorithm audit: Scoring the algorithms that score us. *Big Data & Society* 8, 1 (2021), 2053951720983865.

[8] Amit K. Chopra and Munindar P. SIngh. 2018. Sociotechnical Systems and Ethics in the Large. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (New Orleans, LA, USA) *(AIES '18)*. Association for Computing Machinery, New York, NY, USA, 48–53. https://doi.org/10.

1145/3278721.3278740
[9] Jennifer Cobbe, Michael Veale, and Jatinder Singh. 2023. Understanding aschccountability in algorithmic supply chains. *arXiv preprint arXiv:2304.14749* (2023).
[10] European Commission. 2021. Proposal for a regulation of the European parliament and of the council: laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
[11] European Commission, Content Directorate-General for Communications Networks, and Technology. 2020. The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment. https://doi.org/10.2759/002360
[12] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*. 797–806.
[13] J. Danaher. 2016. The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology* 29, 3 (2016), 245–268.
[14] Mateusz Dolata, Stefan Feuerriegel, and Gerhard Schwabe. 2022. A sociotechnical view of algorithmic fairness. *Information Systems Journal* 32, 4 (2022), 754–818.
[15] Content European Commission, Directorate-General for Communications Networks and Technology. 2019. Ethics Guidelines for Trustworthy Artificial Intelligence. https://doi.org/10.2759/346720
[16] Tracy L Fass, Kirk Heilbrun, David DeMatteo, and Ralph Fretz. 2008. The LSI-R and the COMPAS: Validation data on two risk-needs tools. *Criminal Justice and Behavior* 35, 9 (2008), 1095–1108.
[17] Isabelle Fest, Maranke Wieringa, and Ben Wagner. 2022. Paper vs. practice: How legal and ethical frameworks influence public sector data professionals in the Netherlands. *Patterns* 3, 10 (2022), 100604.
[18] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, and More Authors. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines: journal for artificial intelligence, philosophy and cognitive sciences* 28, 4 (2018), 689–707. https://doi.org/10.1007/s11023-018-9482-5
[19] Jade S Franklin, Karan Bhanot, Mohamed Ghalwash, Kristin P Bennett, Jamie McCusker, and Deborah L McGuinness. 2022. An Ontology for Fairness Metrics. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. 265–275.
[20] Sorelle A Friedler, Carlos Scheidegger, Suresh Venkatasubramanian, Sonam Choudhary, Evan P Hamilton, and Derek Roth. 2019. A comparative study of fairness-enhancing interventions in machine learning. In *Proceedings of the conference on fairness, accountability, and transparency*. 329–338.
[21] L. A. Fujii. 2018. *Interviewing in Social Science Research, A Relational Approach*. Routledge.
[22] De Goede, Bosma, and Pallister-Wilkins. 2019. *Secrecy and Methods in Security Research A Guide to Qualitative Fieldwork*. Routledge.
[23] Golpayegani, Harshvardhan, and Lewis. 2022. *AIRO: An Ontology for Representing AI Risks Based on the Proposed EU AI Act and ISO Risk Management Standards*. ResearchGate.
[24] Nicola Guarino, Daniel Oberle, and Steffen Staab. 2009. What is an ontology? *Handbook on ontologies* (2009), 1–17.
[25] Mark Haakman, Luís Cruz, Hennie Huijgens, and Arie van Deursen. 2020. AI lifecycle models need to be revised. an exploratory study in fintech. *arXiv preprint arXiv:2010.02716* (2020).
[26] Hoekstra, Chideock, and Van Veenstra. 2021. TNO Rapportage Quickscan AI in the Publieke sector II. https://www.rijksoverheid.nl/documenten/rapporten/2021/05/20/quickscan-ai-in-publieke-dienstverlening-ii
[27] Naja Holten Møller, Irina Shklovski, and Thomas T Hildebrandt. 2020. Shifting concepts of value: Designing algorithmic decision-support systems for public services. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–12.
[28] Eric Jonk and Deniz Iren. 2021. Governance and Communication of Algorithmic Decision Making: A Case Study on Public Sector. In *2021 IEEE 23rd Conference on Business Informatics (CBI)*, Vol. 1. IEEE, 151–160.
[29] Pratyusha Kalluri. 2020. Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature* 583, 7815 (July 2020), 169–169. https://doi.org/10.1038/d41586-020-02003-2
[30] Bruno Latour. 1992. Where are the missing masses? The sociology of a few mundane artifacts. *Shaping technology/building society: Studies in sociotechnical change* 1 (1992), 225–258.
[31] Bruno Latour. 1994. On technical mediation. *Common knowledge* 3, 2 (1994).
[32] Bruno Latour. 1999. On recalling ANT. *The sociological review* 47, 1_suppl (1999), 15–25.
[33] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, et al. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–35.
[34] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)* 54, 6 (2021), 1–35.
[35] N. Noy and B. McGuinness. 2001. Ontology development 101: A guide to creating your first ontology. https://protege.stanford.edu/publications/ontology_development/ontology101.pdf
[36] van de Poel and Royakkers. 2011. *Ethics, technology, and engineering : an introduction*. Wiley-Blackwell, United States.
[37] Kit T Rodolfa, Hemank Lamba, and Rayid Ghani. 2021. Empirical observation of negligible fairness–accuracy trade-offs in machine learning for public policy. *Nature Machine Intelligence* 3, 10 (2021), 896–904.

[38] Kit T. Rodolfa, Erika Salomon, Lauren Haynes, Iván Higuera Mendieta, Jamie Larson, and Rayid Ghani. 2020. Case study: predictive fairness to reduce misdemeanor recidivism through social service interventions. In *FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27-30, 2020*, Mireille Hildebrandt, Carlos Castillo, L. Elisa Celis, Salvatore Ruggieri, Linnet Taylor, and Gabriela Zanfir-Fortuna (Eds.). ACM, 142–153. https://doi.org/10.1145/3351095.3372863

[39] Günter Ropohl. 1999. Philosophy of socio-technical systems. *Society for Philosophy and Technology Quarterly Electronic Journal* 4, 3 (1999), 186–194.

[40] J. Saldaña. 2013. *The coding manual for qualitative researchers.* SAGE.

[41] Pedro Saleiro, Benedict Kuester, Loren Hinkson, Jesse London, Abby Stevens, Ari Anisfeld, Kit T. Rodolfa, and Rayid Ghani. 2018. Aequitas: A Bias and Fairness Audit Toolkit. https://doi.org/10.48550/ARXIV.1811.05577

[42] Devansh Saxena, Karla Badillo-Urquiola, Pamela J Wisniewski, and Shion Guha. 2021. A framework of high-stakes algorithmic decision-making for the public sector developed through a case study of child-welfare. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–41.

[43] Lotje Siffels, David van den Berg, Mirko Tobias Schäfer, and Iris Muis. 2022. Public Values and Technological Change: Mapping how municipalities grapple with data ethics. *New Perspectives in Critical Data Studies* (2022), 243.

[44] Giovanni Sileno, Alexander Boer, Geoff Gordon, and Bernhard Rieder. 2021. Like circles in the water: Responsibility as a system-level function. In *AI Approaches to the Complexity of Legal Systems XI-XII: AICOL International Workshops 2018 and 2020: AICOL-XI@ JURIX 2018, AICOL-XII@ JURIX 2020, XAILA@ JURIX 2020, Revised Selected Papers XII.* Springer, 198–211.

[45] J. Spierings and S. van der Waal. 2020. Algoritme: de mens in de machine - Casusonderzoek naar de toepasbaarheid van richtlijnen voor algoritmen. https://waag.org/sites/waag/files/2020-05/Casusonderzoek_Richtlijnen_Algoritme_de_mens_in_de_machine.pdf

[46] Logan Stapleton, Devansh Saxena, Anna Kawakami, Tonya Nguyen, Asbjørn Ammitzbøll Flügge, Motahhare Eslami, Naja Holten Møller, Min Kyung Lee, Shion Guha, Kenneth Holstein, et al. 2022. Who Has an Interest in "Public Interest Technology"?: Critical Questions for Working with Local Governments & Impacted Communities. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*. 282–286.

[47] Anselm L Strauss. 1987. *Qualitative analysis for social scientists.* Cambridge university press.

[48] Harini Suresh and John V. Guttag. 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In *EAAMO 2021: ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, Virtual Event, USA, October 5 - 9, 2021.* ACM, 17:1–17:9. https://doi.org/10.1145/3465416.3483305

[49] Damian A Tamburri, Willem-Jan Van Den Heuvel, and Martin Garriga. 2020. Dataops for societal intelligence: a data pipeline for labor market skills extraction and matching. In *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI).* IEEE, 391–394.

[50] Willem van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. 2011. Design and use of the Simple Event Model (SEM). *Web Semantics: Science, Services and Agents on the World Wide Web* 9 (2011), 128–136. https://doi.org/10.1093/oxfordhb/9780199226443.003.0009

[51] A.F.E. Van Veenstra, S. Djafari, F. Grommé, B. Kotterink, and R.F.W. Baartmans. 2019. Quickscan AI in the Publieke dienstverlening. http://resolver.tudelft.nl/uuid:be7417ac-7829-454c-9eb8-687d89c92dce

[52] Maranke Wieringa. 2020. What to Account for When Accounting for Algorithms: A Systematic Literature Review on Algorithmic Accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20).* Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3351095.3372833

[53] Ben Williamson. 2016. Digital education governance: data visualization, predictive analytics, and 'real-time'policy instruments. *Journal of education policy* 31, 2 (2016), 123–141.

Table 4. Questionnaire

Questions	Notes
Institution / Department	Name of entity/department
What is your (team's) role?	Description team/staff involved brief
Who do you work with (directly)?	
Domain and topic of use case	
Start and end date	
What type of system is being developed for the use case?	Intended use/aim
What is the goal of the system?	
External partners developing technology for the use case (if any)	
Who are the (end) users - are they directly involved in the development process?	
II. Development process	
Could you guide us through the process of development by mapping out phases - and specific actions in each phase?	
Could you guide us through the decisions made about the system and by whom?	e.g. involvement in deciding on: the goal of the system | design of the system (metrics/labeling/test/training/error) | evaluation of system | monitoring | deploying system
What kind of decisions do you and your team make? (could you give an example)	
What input is needed /do you use to make decisions as a reference point	e.g. handbook, training, expert group?
What kinds of exchanges are needed in your decision process?	
Are there other teams or (external) stakeholders involved in the decision-making process?	
How do you support the decision process of your collaborators with your output?	
III. Considerations	
How can the development process be improved for the following, from your perspective?	eg. information exchange, role division, handling error rates and biases, handling risks, responsibilities
What are the most difficult challenges and risks of failures for the system?	
How are these challenges and risks measured assessed and monitored?	
What information is needed (by whom) to handle these challenges and risks?	
Who is consulted for this information?	
What is your role in the process of addressing challenges and risks?	
Could any issues occur that might halt the development process, (if so could you give an example of how are these go/no-go decisions determined?)	
In real-life applications, could there be specific risks or negative impacts for individuals or social groups?	
Is error analysis / Bias analysis performed for negative impacts (if so how is this done and could you give an example)?	
Once the algorithm is deployed in practice, what kind of human oversight is available to control for error, bias, or negative impacts?	
What procedures and recourses, if any, are available for addressing the negative impacts of the system?	
Do you have access to explanations or training on the risks for individuals and social groups, e.g., from your colleagues or from external experts?	
IV. Follow-up questions	
Who is in charge / responsible for mitigating measures on respecting privacy and data protection? For instance, is there a valid legal basis for processing personal data?	
Are there cybersecurity or privacy-preserving measures deployed to preserve privacy and data security?	
If no challenges (or very few) concerns are mentioned in 4, provide a scenario?	e.g. complaints about the output; security breaches; what if the training set is not representative, high error rates