Journal of Responsible Technology 7–8 (2021) 100015

Contents lists available at ScienceDirect

Journal of Responsible Technology

journal homepage: www.sciencedirect.com/journal/journal-of-responsible-technology

The agency of the forum: Mechanisms for algorithmic accountability through the lens of agency

Florian Cech *

Centre for Informatics and Society, TU Wien, Vienna 1040, Austria

ARTICLE INFO

Keywords:
Algorithmic accountability
Agency
Transparency

ABSTRACT

The wicked challenge of designing accountability measures aimed at improving algorithmic accountability demands human-centered approaches. Based on one of the most common definitions of accountability as the relationship between an actor and a forum, this article presents an analytic lens in the form of actor and forum agency, through which the accountability process can be analysed. Two case studies - the Austrian Public Employment Service's AMAS system and the EnerCoach energy accounting system, serve as examples to an analysis of accountability based on the agency of the stakeholders. Developed through the comparison of the two systems, the Algorithmic Accountability Agency Framework (A^3 framework) aimed at supporting the analysis and the improvement of agency throughout the four steps of the accountability process is presented and discussed.

1. Introduction

As current advances in artificial intelligence (AI) and machine learning (ML) prove to be strong drivers of the Digital Transformation of society, the widespread adoption of automated decision making (ADM) and decision support systems raises more and more urgent calls for increased accountability of socio-technical systems. Following the global trend towards evidence-based policy and decision making,[1] these systems are often presented by their creators with the promise of increased efficiency and fairness (Elish and Boyd, 2018; O'Neil, 2016). While positive examples of how the use of algorithmic technologies can improve efficiency or even reduce human bias do exist (Skirpan and Gorelick, 2017), a mounting evidence that betrays these promises suggests the opposite: systems like Amazon's AI recruitment tool (Dastin, 2018), the COMPAS recidivism risk assessment system (Brennan et al., 2008) or the Austrian Public Employment Service's' AMAS tool (Allhutter et al., 2020a) all carry the risk of both perpetuating pre-existing and introducing technical and emergent biases into administrative processes - three separate types of bias that were identified as early as 1996 by Nissenbaum (1996) in her seminal work on accountability in a computerized society. Furthermore, these systems may even result in reduced efficiency given the additional tasks required of 'human--on-the-loop' actors (Wagner, 2019) tasked with monitoring and correcting problematic decisions or outcomes. Given the fact that these algorithmic systems represent, in the words of Bruno Latour, "society made durable" (Latour, 1990), the real and present danger of fundamentally biased and unfair systems that are becoming more and more ubiquitous and have a long-lasting impact on human lives urgently demands solutions to hold these systems accountable. Beyond ADM, other types of algorithmic governance such as sustainability and eco-feedback tools (Cech, 2021; Froehlich et al., 2010; Strengers, 2011), while not necessarily affected by issues of bias and discrimination, can have equally strong impacts on society and may require similar levels of accountability.

While political institutions such as the European Commission, among others, recognize these issues and put forth proposals - for instance, the 'Artificial Intelligence Act' (European Commission, 2021) - scholars have been quick to point to the planned regulation's shortcomings when it comes to providing the necessary safeguards for citizens affected by algorithmic decision making, particularly when the system in question is embedded in governmental or public bodies' decision making processes (Fink, 2021). As it stands, accountability regulation often falls woefully short when it comes to implementing concrete procedural, social, or technical measures. To give an illustrative example, the European Union's GDPR guarantees a "right to explanation" for algorithmic decisions, but - as Wachter et al. lament - fail to define "*[...] what type of*

* Corresponding author.
E-mail address: florian.cech@tuwien.ac.at.

[1] See Greenhalgh and Russell (2009) for a well-rounded critique of positivist paradigms in policymaking.

https://doi.org/10.1016/j.jrt.2021.100015
Received 29 June 2021; Received in revised form 1 October 2021; Accepted 4 October 2021
Available online 20 October 2021

explanation is intended or what purpose it should serve" (Wachter et al., 2018, p.862), which forces both subjects of algorithmic decision making and data controllers to interpret what types of explanations are necessary and appropriate in specific cases. Beyond specific procedural or technical measures, only few social measures such as educational resources to raise the algorithmic literacy in the populace have been implemented, let alone codified in law as requirements for the use of algorithmic systems. This may be due to the fact that, from a scientific standpoint, many questions as to how larger societal measures could be designed to address these issues are still unknown and the focus of active research (Hargittai et al., 2020; Lomborg and Kapsch, 2019).

One of the most prominent reasons for these issues lies within the *wicked* (Rittel and Webber, 1973) nature of the problem: algorithmic accountability is a multi-faceted challenge that, depending on the context, lacks both a *"definitive formulation"* and *"ultimate [...] test of a solution"* and, at best, can be approached to reach a "satisficing" solution, as (Fitzpatrick, 2003, p.4) describes. Consequentially, any legislative proposal to improve this situation can either only make very general recommendations or prescribe specific measures that will work in some, but not all cases. This is not a challenge unique to algorithmic accountability; many other social issues and wicked problems exist that require nuanced approaches within the law. What makes algorithmic accountability particularly difficult to address solely from a legal standpoint are, among others, two key points: the (sometimes inherent) *complexity* of the underlying systems on the one hand, and the *rapid deployment* and *wide-ranging effects* a single system can have on vast numbers of people. On the complexity side, the discrepancy in knowledge and understanding between legal experts and domain experts in computer science makes legislation as challenging a process as, for example, regulation for aircraft safety (Roelen and Klompstra, 2012). Contrary to aviation technologies, though, algorithmic systems can be created, deployed, adapted, scaled and even disbanded again literally in time-spans as small as seconds. Once deployed, a system written by a handful of people can affect hundreds of millions, which is the foundation many of today's tech startups rest on. Consequentially, this both offers a fantastic potential for innovation on the one hand, and, on the other hand, poses serious risks in case something goes wrong. To legislate for such a complex and fast-moving target is, thus, particularly challenging.

These regulatory challenges notwithstanding, the wicked problem of accountability expands to a more fundamental and concrete level. Even under the assumption that, for a given context of an application, concrete laws and guidelines exist to make systems accountable, it is a matter of spirited debate among the scientific community *how* exactly to design systems and procedures to implement such laws in a satisfactory manner. To approach the problem, much attention (Burrell, 2016; Janssen et al., 2017; Pasquale, 2015; Rader et al., 2018) has been given towards the question of *explainability* and *transparency* as necessary preconditions for accountable systems, both in the sense of *system transparency* and *post-hoc explanations* (Mittelstadt et al., 2019) for specific outputs of such systems. Particularly for AI/ML applications, explainability is a fundamental challenge, given the vast amount of data these systems require to train and operate and the opaque steps of statistical correlation leading to the outcomes. However, even less sophisticated systems may present a significant challenge to transparency and explainability, depending on *who* is affected by the system, *where* it is being deployed, and *how* it is being operationalized. While there is little doubt that purely black-boxed systems (Pasquale, 2015) can hardly satisfy reasonable standards of accountability, transparency alone does not guarantee accountability either (Ananny and Crawford, 2016), and, as a problem, shares many of accountability's *wicked* characteristics (Cech, 2021).

Based on the challenges outlined above, it stands to reason to conceptualise algorithmic accountability not just from a legal standpoint, but more fundamentally as a general *process* that can play out in many contexts - one of them the legal system, but more often between the people affected by algorithmic systems and decisions and those implementing, operating and maintaining these systems. In this article, I suggest an approach to accountable algorithmic systems that is applicable in a wide range of scenarios, based on the widely-used definition of accountability by Bovens (2007) relating *actor, forum* and *account*. Within this definition and for the purpose of this analysis, people affected by algorithmic decision systems represent the *forum*, whereas those operating and maintaining the systems represent the *actors*. While this serves as a specific example, the proposed approach can be adopted for other definitions of *actor* and *forum* as well, which I address in Section 7.

Supported by two examples of algorithmic systems - the EnerCoach eco-feedback tool as a successful, and the Austrian Public Employment Service's AMAS tool as a problematic one - I will focus on accountability through the lens of *agency* of the involved *actors* and, particularly, of the *forum* in the above mentioned definition. Drawing a parallel to the methodological approach of participatory design, the two systems will serve as illustrative examples of (1) how a lack of meaningful stakeholder involvement can lead to reduced accountability and transparency of the system and (2) how empowering design methodologies can improve the accountability of a system through the co-creation of transparency and accountability measures. Comparing the two case studies yields learnings derived from their commonalities and differences. To help future efforts, I then propose the Algorithmic Accountability Agency Framework (A^3 framework) that provides a set of structured guidelines to facilitate the assessment and improvement of the accountability of algorithmic systems. Finally, I conclude with a summary and an outlook towards future work to help develop further accountability measures for algorithmic systems.

2. The actor, the forum and the account

One of the most widely cited definitions of accountability (perhaps except for the one provided by Giddens (1986) as part of structuration theory) in the context of computing is Bovens', who describes accountability as

> *"[...] a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor may face consequences."* (Bovens, 2007, p. 1)

While this definition was coined at a time that algorithmic accountability was not yet the widely recognized issue it has become today and stems specifically from the field of accountability studies and governance, his characterization allows for a broad application of the term to other fields as well. As Wieringa (2019) notes, regulatory actions such as the European Union's General Data Protection Regulation (GDPR) and open government initiatives (Fink, 2017) have put the term 'algorithmic accountability' front and center, leading many scholars working on similar topics to lean on Boven's definition.

Bovens explicates different types of accountability depending on the *levels* of actors (*individual, hierarchical, collective* and *corporate*), *roles* of actors as *decision makers, developers* and *users*, as well as *type* of the forum (*political, legal, administrative, professional*, and *social accountability*).[2] Within these categories, most kinds of accountability relationships related to algorithmic systems can be described: from the jobseeker affected by a risk assessment algorithm demanding to see the reasoning for their score, to civil society watchdogs critiquing the design and operationalization of an algorithmic system used in predictive policing, to name just two examples, the definition can be used to model a variety of accountability processes with vastly different requirements, actors

[2] See Wieringa (2019) for a more detailed characterization of these categories.

and potential outcomes.

A closer look at all these processes then reveals complex levels of *agency* required for different accountability interactions. *Agency* as a theoretical concept has been characterized as 'elusive' and 'vague', yet 'resonant', and gets associated by theorists with a variety of other concepts, including motivation, will, purposiveness, intentionality and choice, as Emirbayer and Mische (1998, p. 1) explicate. For the purpose of this article, a much simpler definition suffices: the ability to act by one's own volition and choice. Throughout the three phases of the accountability process as proposed by Brandsma and Schillemans (2013) - *information gathering, deliberation*, and *consequences* - we can identify four separate types of *agency* required of either the forum or the actor: the ability to *ask for information*, to *provide answers*, to *impose consequences*, and to *effect change*. Fig. 1 shows a visualization of these 4 types of agency required by the forum and the actor respectively.

The first step entails the agency of the forum to pose their questions. Practically, this requires that the given forum - be it an individual, a civil society watchdog, journalists, legal or executive entities or society at large - is aware of (1) the existence of the system, (2) the identity of the actor (e.g., the developers, maintainers or product owners of the system) and (3) have means of communicating their inquiries. In many cases, this agency may already be limited by the forum's limited knowledge about the system or its existence - a problem that falls clearly into the domain of *transparency*, but also relates to the *algorithmic literacy* of the forum. Even if the forum has the required understanding and knowledge of the system, identifying the actor to pose the questions too can be an impossible task, which Nissenbaum (1996) described as the "many--hands problem": with the increased complexity of these systems, the number of involved actors rises as well, both in terms of individual humans and organizational entities. For example, a single system like Austria's AMS Algorithm (Allhutter et al., 2020a) may be conceptualized by one entity, designed and developed by another, audited by a third and maintained by a fourth, allowing each of those actors to refuse responsibility and send any inquiries into a never ending circle of referrals.

However, the nature of such referrals may not be malicious but stem from limited agency on the side of the actor in the second step as well. Legal requirements and limitations, unclear internal responsibilities, a lack of resources, or simply lack of knowledge may also prohibit even the most well-meaning actors from fulfilling their side of the accountability relationship in the information-gathering phase. Mismatched questions or misunderstandings on the side of the forum, as well as the complexity and nature of the system itself, may also impact a willing actor's ability to provide the answers requested. Taking the example of facial recognition and its well-documented bias against black and brown faces (Buolamwini and Gebru, 2018), an individual falsely identified by police through facial recognition may well be able to file a complaint against the police and demand an explanation. However, even if all involved actors (such as the police themselves as well as the company responsible for developing and maintaining the recognition software and supplying the training data) were entirely willing to cooperate, they might not be able to conclusively answer the question 'Why was this specific person falsely identified?'. This is a direct consequence of the inherent complexities of machine learning and image recognition technologies. As Mittelstadt et al. summarize, many state-of-the-art machine learning methods like Artificial Neural Networks / Deep Learning rely heavily on black-box functions that operate on *"[...] an internal state composed of millions of interdependent values"* (Mittelstadt et al., 2019, p.1). Consequentially, even domain experts can only deliver explanations that are extremely technical, such as certain gradients crossing a threshold that classifies part of an image as a match. These explanations may be technically correct, but entirely useless for the individual demanding an explanation for an algorithmic decision that affected them.

The third step - imposing consequences - is where the forum's agency is most obviously limited by hegemonial power imbalances. In the few cases where algorithmic bias can be proven to be the source of discrimination in the legal sense, legal action may be the most crucial type of agency the forum has. Short of strong legal requirements and accountability regulation that expands this agency to the myriad of legal grey areas concerning algorithmic governance, the forum's agency to impose consequences may be limited to publicizing their findings, organizing against the actors, or otherwise starting a public discourse about the system. While creating grass roots movements, public outcry, and media attention can certainly be a powerful type of agency (Joyce et al., 2020) and have had significant success in the past, they are also limited to a forum that is not a single individual, but rather a collective (in the sense of *political* or *social* accountability as described by Bovens, 2007).

The fourth step - effecting change - is not explicitly part of Boven's

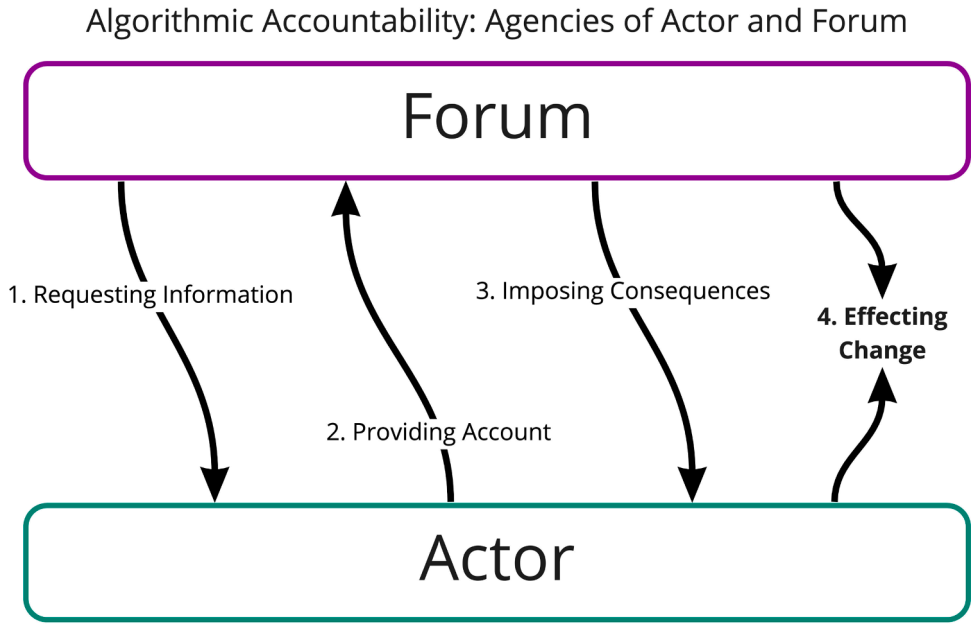

Fig. 1. Visualization of the accountability process.

definition but follows implicitly from the intention of making algorithmic systems more accountable: After all, any consequences imposed by the forum that do not at least carry the potential to lead to any change are meaningless. There may be fundamental disagreements between the forum and the actor on whether or not the conduct held to account is justified, as outlined by Binns (2017) with the example of automated credit scoring:

> *"For instance, the bank might justify their decision by reference to the prior successes of the machine learning techniques they used to train their system; or the scientific rigour involved in the development of their psychometric test used to derive the credit score; or, more fancifully, divine providence. The loan applicant might reject such justifications for various reasons; for instance, on account of their being sceptical about the machine learning technique, the scientific method, or the existence of an interventionist God, respectively."*
> (Binns, 2017, p. 2)

These potentially irreconcilable *"[d]ifferences of opinion about the normative standards"* (Binns, 2017, p. 6) notwithstanding, a lack of agency on the side of the actor to effect change - from, in the simplest form, the adaptation of the algorithmic system to avoid the conduct in question, or in more extreme cases, discontinuing the use of the system altogether - would *a priori* negate any meaningful accountability process. This type of agency may seem impossible to guarantee after a system has been deployed: adapting any given system to a nigh-infinite number of edge cases that may, in some cases, negate the use of the system altogether can be as futile a proposition as expecting that systems developed by expending significant resources would easily be abandoned due to unjustifiable conduct in a minority of cases. Subsequently, this type of agency resides in the complex domain of political decision-making, algorithmic governance and moral responsibility, and will require a broader societal discourse than can be covered by the accountability process alone.

The outline of the complex and manifold different types of agency required for the actor and the forum alike given above illustrates the fragility of the accountability process. If just one of the entities in these four steps lacks the ability to act, the accountability process is doomed to fail without a satisfying, or even 'satisficing' (Fitzpatrick, 2003), solution. Consequentially, any technical, social, or procedural measures designed to improve accountability must preserve or enable these kinds of agency first and foremost, lest they can easily be disregarded due to their ineffectiveness.

The following two sections illustrate the relationship between agency and accountability through two case studies and outline the potential for actor and forum agency (or lack thereof) in the context of two specific algorithmic systems in the form of a comparative case study (Knight, 2001).

3. The AMS algorithm: stakeholders, design process and the lack of agency

In 2018, the Austrian Public Employment Service (AMS) announced plans to introduce a statistical profiling system named AMAS[3] to classify jobseekers into three categories: those with high, medium or low chances of reintegration into the labour market within a short (7 months) and a long (2 years) time frame. As Allhutter et al. (2020a) outline, the stated goal of the system was threefold: to reduce unconscious human bias of caseworkers, to increase the efficiency of the caseworker - jobseeker interactions, and to increase the effectiveness of resource-intensive measures (such as subsidized employment, retraining or further educational measures) granted by the AMS. At the core of the system, a comparatively simple algorithmic system compared jobseekers based on a number of variables with previous jobseekers with the same or similar data points (including age, gender, previous employment history, health impairments, and duties of care as documented by Allhutter et al. (2020a, p. 6)). The system used this comparison to calculate the ratio of those previous jobseekers that had managed to find and keep a job within the given time frames to those that had not, and used that ratio to classify jobseekers into one of the three categories listed above. The AMS algorithm was meant to be integrated as an assistive system into the normal process of consultation between jobseekers and AMS workers and - depending on the classification - restrict which measures would be possibly available to the jobseekers. While the AMS claimed that the nature of the system was purely assistive, and that the final decision would always be made by a human, the Austrian Data Protection Agency forbade its use based mainly on the assessment that there were insufficient measures to guarantee AMS workers would not routinely follow the system's recommendation[4] - a practice that would violate Austrian laws on automated profiling.

A broad and critical public response from scholars and advocacy groups followed the first announcement of plans to use the AMAS system. Among the major points of critique were the potential of the system to discriminate against jobseekers based on very limited data, intersectional bias against particular populaces, and the danger of creating a feedback loop that would cement pre-existing bias and discrimination on the labour market. A general lack of (system) transparency detailing how the system actually worked and would affect jobseekers, as well as the lack of socio-technical evaluation or technology assessment studies prior to implementing the system was also noted by scholars (Allhutter et al., 2020a; 2020b; Wagner et al., 2020).

A more detailed look at a prototypical accountability processes reveals a number of issues that require further consideration. The first step for such an analysis is identifying the actors and the forum. For the purpose of this illustrative example, a simplified version would look like this: the *AMS caseworker*, as part of their consulting process with the jobseeker and as the 'user' of the system, would be the *actor* that the *jobseeker*, as the *forum*, demands justification from. At this point, identifying the caseworker as the actor (in lieu of, for example, the AMS as a whole or the agency that developed the system) is but one way of modelling the accountability process. Most current approaches to the accountability process consider not the level of specific human actors, but rather the organizational or policy levels. While this is certainly a valid approach that has its merits, as explicated in more detail in Section 6.1, it also does not represent a human-centered approach to accountability. To realize the potential of the agency-based lens to model accountability even on this micro-level of human-to-human interactions, the prototypical accountability process in this case study focuses directly on the case worker and job seeker. This approach reflects that, as a general matter of policy, the AMS encourages its case workers to explain to the jobseekers, upon request, how the system works and how it arrives at the classification presented, and act as a representative for the larger organisation as part of this accountability process.

Following the analytic lens of the 4 steps outlined in Section 2, the jobseeker would first require the agency to ask the AMS worker to justify their conduct - in this case, their decision to follow or overrule the AMAS system's classification. This requires the jobseeker to (1) be aware of the existence of the algorithm, its general purpose and specific output and (2) suspect a false or improper assessment by the system. Given the issues of transparency and the fact that not all jobseekers can be expected to have the required knowledge and algorithmic literacy to fulfil both requirements, at least some of the people affected by the system already do not have the agency to take the first step in this accountability

[3] orig. "Arbeitsmarkt-Assistenz-System", or "Labour Market Assistance System" (translated by the author).

[4] This ruling was subsequently suspended by the Austrian Federal Administrative Court; an appeals process is ongoing.

process. Under the assumption that the jobseeker in question does utilize their agency and demand justification for the result of the decision process shared between AMS worker and AMAS system, the AMS worker only has very limited tools and resources at their disposal. The AMAS system does provide some explanations for the classification and score in the form of pre-formulated, modular texts presented to the caseworkers, which outline - in a very generalized form - the major impact factors leading to the positive or negative assessment. For instance, one sentence explaining a 'challenging' factor reads

> *"Your were only employed over limited periods of time during the last few years."*[5] (Allhutter et al., 2020b, p. 72)

While this does constitute a form of account for the results of the system, it does so in an extremely condensed and limited form. Neither do these texts clearly state what specific variable constellation triggered this negative assessment, nor does it clarify the impact this particular variable had on the overall assessment. Furthermore, the system does not provide these explanatory texts for job seekers with an 'incomplete employment history'.[6] Consequently, the caseworker's agency is limited to providing these very condensed explanations for roughly two thirds of all jobseekers (about 31% of jobseekers at the beginning of their case only have incomplete employment histories (Allhutter et al., 2020b, p. 31)). On top of these limitations in terms of available tools, the AMS workers also operate under tight time limitations: As Penz et al. (2017) note, caseworkers on average only have 15–30 min to spend talking to their clients - a number that, in light of the surge of unemployment following the COVID-19 pandemic has inevitably been reduced even further. Given the numerous other bureaucratic tasks caseworkers have to accomplish in that time frame, it stands to reason that any spending time on giving a more detailed account for the AMAS system's conduct is simply not possible in most cases.

For the remaining jobseekers (i.e., those who could utilize their agency to ask for justification, and of those the ones that could receive at least some kind of response), their next step would be the deliberation of the answer and the question of imposing consequences. Unfortunately, the AMAS system itself offers no technical measure to facilitate such a request. The only option is the procedural assurance given by the AMS itself that caseworkers would discuss the results with the jobseeker and try to find an agreement. If that can not be found, the jobseeker's only option is to file a complaint with an 'ombudsman' within the AMS structure and go through an arbitration process, in which they have precious little information to support their suspicions that they may have been wrongly classified. Given that the system would be used on roughly half a million jobseekers every year and, on average, claims an accuracy of 85%, that would leave roughly 75.000 people vulnerable to a misclassification that may impact their future job chances, without a reasonable way of challenging their results.

Finally, under these adverse circumstances, the question for effecting sustainable change remains unanswered. Within the confines of the example given above, where the actor is defined only as the caseworker, there is no plausible way for the caseworker to effect institutional change that would lead to an adaptation of the system. This limitation notwithstanding, the operationalized of the system as part of the consultation process does leave the caseworker the agency to routinely contradict the system, albeit at the cost of efficiency, since the AMS requires caseworkers to provide a written explanation any time they choose to change the category assignment of a jobseeker.

An in-depth qualitative textual analysis of the roughly 135 internal protocols, requirements documentation, meeting/workshop minutes, and internal educational materials used as the basis for the work by Allhutter et al. (2020b)[7] reveals the reasoning for and the development process of the only part of the project concerned with accountability: the textual explanations given to jobseekers with complete data only. While the need for post-hoc explanations (Mittelstadt et al., 2019) of the results was brought up early in the development process, the decision on what form these descriptions should take seems to be largely influenced by technical limitations and requirements. While early protocols reveal a discussion about 'graphical visualizations' as explanations, later protocols only focus on a textual representation and show that the design and implementation were by and large aimed at the AMS workers rather than at the job seekers. Furthermore, an internal note from the IT department of the AMS reveals that - for technical reasons of how the interface of the internal software was designed - texts would have to be limited to 255 characters. A workshop on the specific explanatory text modules was held solely with the AMS workers and did not include any jobseekers, and only happened after the formal and technical requirements were already defined. Finally, the texts themselves were based on suggestions by the contractor implementing the AMAS models and algorithms and frequently seemed to have been based on assumptions about the causality between correlating observations: for instance, one internal document lists the example of a correlation of the age group 50+ with certain results as based on 'job experience' for a positive effect, and the fact that an individual's education 'lies far back in the past'[8] - two causal connections that are not implausible, but simply not supported by empiric evidence disclosed through the algorithm, and certainly not deducible from the regression models describing the correlations.

What this analysis illustrates is a funnel of diminishing agency for both actor and forum from the start to the end of the accountability process, reducing the chance for a successful accountability process with each step. What little measures were implemented to improve the accountability of the system were aimed more at the agency of the actor than the forum, and did not include any jobseekers in their development, neither in a co-design nor in an evaluative fashion. The fact that these measures were considered only after the technical implementation of the system was agreed upon betrays the techno-solutionist nature (Morozov, 2014) of the system in general. The result is an implementation sorely lacking in ensuring the agency of jobseekers to hold the system to account.

4. EnerCoach: participation and the benefits of empowering the forum

The focus of the second illustrative case study - the EnerCoach collaborative energy accounting system[9] - is a web-based algorithmic system currently used by over 700 Swiss communities, and is a fairly complex example of a class of sustainability tools called eco-feedback tools (Cech, 2021; Froehlich et al., 2010; Strengers, 2011). Its main functions are the collection and aggregation of municipal energy data (e.g., energy consumption, costs or CO_2 expenditure), as well as a complex analytic system that generates a variety of graphical reports giving a normative assessment on a community's sustainability performance. These reports are then used by energy auditors to certify communities as part of the Swiss EnergyCity[10] program, as well as the European Energy Awards (EEA) program.[11]

While the EnerCoach system does not employ machine learning or

[5] orig. "Sie waren nur über kurz Zeiträume in den letzten Jahren beschäftigt"; translated by the author.

[6] orig. "unvollständiger Erwerbshistorie" (Allhutter et al., 2020b, p. 61); translated by the author.

[7] For a complete list of included materials see Allhutter et al. (2020b, pp. 111-116)

[8] orig. "Beispiel 'Alter 50+' - stärkend im Sinne von Berufserfahrung, hemmend im Sinne von weit zurückliegender Berufsausbildung"; translated by the author

[9] See https://www.local-energy.swiss/infobox/energiebuchhaltung.html

[10] See https://www.energiestadt.ch/

[11] See http://european-energy-award.org/

AI-related technologies to calculate its results, the complexity of energy accounting and reporting leads to a significant challenge in terms of making the system transparent to its various stakeholders, as a recent study shows (Cech, 2021). Among others, these stakeholders include domain experts such as energy consultants, auditors and the EnerCoach hotline,[12] but also end-users tasked mainly with data entry in the form of municipal workers or facility managers. Finally, the "EnerCoach working group [...] serve as a steering committee and define goals for the further development of the tool in general and calculation policies in particular" (Cech, 2021, p.3). As noted in the study, ensuring the system's accountability regarding the correctness of the results hinges mainly on issues of system transparency and post-hoc explainability. A typical accountability process for this system would entail the forum - an energy consultant, auditor or community user - trying to determine whether an implausible report result stems from incorrect data entered by the community or an error in the calculation process, and the actor - the system developers, or as a mediating instance, the EnerCoach hotline - trying to justify the results given by the system. Given the fact that the report results may significantly impact a community's ability to reach a sustainability certification, errors in the system have the potential to cause adverse political or financial consequences for the community, making the accountability of the system and the user's trust in it an important factor in its use. Furthermore, while there are no immediately applicable legal requirements for the operators and designers of the system to provide an account for its correctness, an individual incentive for accountability exists in the in sense of *reputational concerns*. Buhmann et al. (2020) identify the *"a pragmatic necessity and a normative obligation for organizations to take part in fora that allow for an inclusive debate on the workings and ramifications of the algorithms that they employ."* (Buhmann et al., 2020, p.272) For the case of EnerCoach, this pragmatic necessity exists to encourage the ensure the further use of the system in order to work towards the greater goal of sustainability and, in a larger sense, the fight against climate change: what is at stake in this case is not legal compliance, but the community's trust in the correctness of the system and its value for sustainable community energy practices.

An analysis of this exemplary accountability process along the 4 steps of agency reveals the following challenges. First and foremost, the forum must come to the conclusion that a given report result seems implausible, and suspect an error in the system. The forum's agency on this aspect is limited by their energy accounting domain knowledge; depending on whether the forum in this case is an energy auditor with a high degree of understanding of the underlying calculations or a community user with a lesser understanding, there may be significant differences in how easily the forum can reach this conclusion. However, as the underlying data for the reports are provided by the forum itself, verification by comparison is possible in the same sense that Sandvig et al. (2014) proposes field experiments to audit black-boxed algorithmic systems for discrimination: trying to enter different data and comparing the changes may reveal already an answer to the demand for justification, and subsequently even include the possibility for change if the error was to be found in the data rather than in the calculations itself. In case this approach does not provide conclusive answers and to escalate this accountability process further, the forum in question would then have to make a support call to the EnerCoach hotline to request help with their particular problem. Their agency to do so is limited only by their knowledge of the hotline's existence; given that the system offers contact addresses and phone numbers for the hotline, that help can be requested free of charge, and that new users go through training workshops (Cech, 2021, p. 3) that include references to the hotline, most users should be able to do so without limitations.

The second step - providing the account and justifying the system's conduct - entails differing levels of agency depending on the identity of the actor. As first point of contact, the hotline is staffed by energy experts that can utilize their experience to assess the correctness of the data entered and manually calculate rough estimates of the results. Their agency is mostly limited by their access to and knowledge of the system's underlying code base; for any inquiries that go beyond verifying the plausibility of the results, they refer the request to the software developers responsible for the implementation of the system. Their agency to provide a justification is only limited by socio-economic factors, but not necessarily by technical ones: while they have access to all underlying code and data, as well as the required knowledge to trace any implausible results to their roots, this process can be resource- and time-intensive, and may be limited by the costs incurred. As the system is currently financed by public funds Cech (2021, p.3), which include a certain number of hours reserved for providing this kind of support, there is at least a minimum of agency to provide a detailed account in for these questions.

The agency to impose consequences and to effect change (steps 3 and 4) is more individual and depends on the identity of the forum and actor. While some energy consultants, auditors and community users may have personal relationships to the EnerCoach working group that could authorize and finance such adaptations or bug fixes, others may simply be administrative personnel working in a municipal office and may not have this kind of access. On the side of the actor, the EnerCoach hotline can escalate and prioritize technical issues that come up more frequently for adaptation with the EnerCoach working group as well. Finally, as the system is undergoing regular maintenance and bug fixing, the development company has the agency - provided they are granted the necessary resources - to address any technical issues and adapt the system according to the requirements resulting from the accountability process.

In assessing this accountability process through the lens of actor and forum agency, the most restricting factor emerging is the transparency of the system. As outlined above, transparency dictates the specificity of the questions posed to the actor, and may in some cases even conclude the accountability process without the need for further action. To address this issue, Cech (2021) chose a participatory design (PD) approach, including relevant stakeholders (energy consultants, the EnerCoach hotline and users) in the process. As they describe their methodology, the participants in this process were tasked with developing transparency measures aimed at supporting the sense-making processes (Lebiere et al., 2013; Wood et al., 2019) involved in tracing implausible report results and improving system transparency for both users and the hotline in general. The result was a set of flow-chart visualizations of the algorithmic processes necessary to create the reports, as well as a number of smaller improvements to the user interface of the tool (such as user-defined data annotations that would be shown alongside the final reports to trace anomalies in the input data to their respective output). While they report on the success of the methodology to design and improve existing transparency measures, an interesting aspect relating to agency goes beyond the specific measures resulting from this process: by involving the affected actors and members of the forum for these accountability processes in EnerCoach, they establish a new kind of agency to effect change, namely to (positively) influence the accountability process itself. The potential of participatory methods to empower users traces back to the roots of this discipline in the worker's movements of the 1960s and 1970s in Germany and Scandinavia (Wagner, 2017, p. 247, citing Kensing and Greenbaum, 2012). In the case of EnerCoach, following a co-design approach with a diverse user group meant that (1) participants were knowledgeable about the system and its issues through their own experience and (2) had a personal stake in the result. Both of these factors contributed to the success of the intervention, and helped avoid an approach driven by techno-solutionism that might have led to other, less appropriate development of technical transparency measures. Furthermore, methodologies such as participatory design or other co-design approaches

[12] A group of energy experts responding to support requests of end-users; see Cech (2021, p. 2) for the detailed stakeholder analysis.

contribute to elevating the stakeholders involved to a more "critical audience", a prerequisite for algorithmic transparency as explained by Kemper and Kolkman (2018).

In summary, the EnerCoach case study can be characterized as a success story for algorithmic transparency and accountability not simply for the concrete measures implemented, but rather for the positive effects on stakeholder agency the methodology of participatory design brought to light.

5. Comparison of the case studies

Juxtaposing the two case studies as illustrative examples, one might conclude that they are not comparable, as they pertain to two different domains of operation, differing algorithmic methodologies and stakeholders, as well as different reasons why accountability represents a value identified as either required or worth pursuing by the system stakeholders. These differences notwithstanding, for both systems, accountability has been identified as a necessity and a complex challenge in terms of the concrete implementation of accountability processes. Upon a closer look through *actor* and *forum agency* as a framing device, the following further comparable aspects emerge in the analysis.

First and foremost, both systems share a certain *technical sophistication* that makes transparency and post-hoc explainability difficult. The AMAS uses a big data approach and calculates statistical averages and ratios for hundreds of thousands of job seekers over an extended period of time. Even if the calculations for a single case in the final system may be trivial ratios between those jobseekers that fulfil certain criteria and those that do not, providing meaningful explanations can become a needle-in-a-haystack problem, and is further limited by GDPR and privacy considerations. The EnerCoach system, on the other hand, starts out with comparably simple data points (i.e., electricity meter X in object A logged a consumption of 10.000 kWh from 1/1/2020 until 12/31/2020), the final reports require a complex interplay of various other data points, such as the building size and location, climate correction factors, and normative design choices such as the way energy efficiency is assessed. Both system's complexity defies giving simple explanations, although the explanatory text snippets provided by the AMAS system are certainly attempting to do so with questionable success.

Secondly, both systems share a *complex assemblage of stakeholders* with varying degrees of algorithmic literacy. Even leaving the system developers and the larger organizational context of the AMS itself out of the equation, the AMS workers themselves may well be domain experts in terms of the regional labour market, but have only a limited understanding of the inner workings of the system as given by the educational materials the AMS supplied them with. The jobseekers, on the other hand, represent a representative cross-section of the Austrian populace, perhaps with a certain bias towards certain professions and educational backgrounds. As such, neither domain knowledge on the labour market nor technical knowledge on the statistical processes utilized by the AMS algorithm can not be expected. Similarly, the different user groups of the EnerCoach system can also range from domain experts to administrative workers with limited domain knowledge, and neither of them necessarily have insight into the inner workings of the system.

Based on these two observations, we can draw a third comparison regarding issues of transparency and post-hoc explainability. Both are socio-technical assemblages (Latour, 2007) that have been found lacking in regards to explaining specific results, either by themselves through purely technological measures, or through human mediation (such as the AMS caseworker or the EnerCoach hotline). Both systems employ certain measures to tackle this challenge: the AMAS system through the explanatory text snippets, and the EnerCoach system through visualizations and small adaptations to the interface. One main difference leading to the success or failure of these measures as posited by this article is the process of developing these measures, and their inherent potential to support or restrict the agency of the forum and actor. The AMAS measure was designed, evaluated, and implemented solely with the actor, i.e., the AMS worker, in mind, and there is no indication that the agency of the forum played any significant role in its design. Even the agency of the AMS worker in the design process is in doubt, since the possibilities of what shape or form this transparency measure would take were limited beforehand by technical and procedural requirements. In contrast, the EnerCoach system's improvements towards transparency were conducted with the stakeholders and affected parties, including actors *and* forum, and thus were designed to improve the agency of both sides throughout the first two steps of the accountability process.

Even though, at face value, both systems seem to be quite different from each other, the framing device of accountability agency allows a useful comparison, and suggests a wider applicability to other systems facing similar challenges of diverse stakeholders with varying degrees of algorithmic literacy and domain knowledge, and complex algorithmic processes defying simple solutions to transparency and explainability.

6. The accountability agency framework

Supported by the comparative analysis of the two case studies, I propose the Algorithmic Accountability Agency Framework (A^3) as an analytic lens to assess and improve an algorithmic system's accountability. For each of the 4 steps visualized in Fig. 1, Table 1 provides a set of guiding questions, which can both help to analyse the status quo of a given algorithmic system as well as provide a starting point for the development of further measures to improve the situation.

To make the best use of the framework, a comprehensive stakeholder analysis of the system in question is required to identify all reasonable actor/forum combinations. For instance, the case study of the AMS algorithm would yield jobseekers themselves, advocacy groups representing them, academics in various disciplines (including computer science, science and technology studies, and law), and governmental entities (such as the Austrian Data Protection Agency) and so forth as various forums. On the other side, besides the AMS workers using the system, potential actors would be the AMS management or the contractor developing the system. While not all possible combinations of forum and actor would result in meaningful accountability processes, it may still be helpful to explore other types of accountability processes beyond the most obvious one. For instance, the interaction between academics studying the system as forum, and the contractor as actor, requires different types of agency and allows for different (asynchronous) channels of communication, as opposed to the interaction between jobseeker and AMS worker, which has to be synchronous and time-constrained.

The primary guiding questions are posed to shine a light on the various restrictions and limitations that result from the system's context of use. While this approach is particularly useful for the analysis of an existing system, many of the restrictions and limitations may not be easily predicted for a system currently being developed. To make the framework useful for these cases as well, the set of secondary questions aims to help guide the design and implementation of measures to improve the agency of the forum and the actor. To illustrate, when looking at the agency of the forum in step one for the EnerCoach system, asking the question "What information or domain knowledge does the forum require to formulate an inquiry?" may lead to the answer that the user needs information the system does not provide, prompting the design of a new feature to improve the users' agency by supplying that information.

While the questions are meant to be supportive of accountability, following the framework and providing answers to all questions does not guarantee a positive accountability process. System-specific issues, as well as the question of disagreement between forum and actor on the validity of the justification (Binns, 2017), as well as the ultimate opportunity to impose consequences and effect change, can not be generalized as easily. What the framework can provide, nonetheless, are the minimum requirements that make any kind of accountability possible

Table 1
A^3 framework guiding questions for each step of the accountability process.

Step	Guiding Questions
1. Requesting Information	**Which factors may restrict the ability of the forum to request information?** How can the forum be made aware of the existence of the system? What information or domain knowledge does the forum require to formulate their inquiry? How can the forum identify the actor? What channels of communication can the forum use to communicate with the actor?
2. Providing Account	**Which factors may restrict the ability of the actor to provide an account?** How can the actor respond to an inquiry? What information or domain knowledge does the actor require to provide an account? What tools does the actor have at their disposal to collate the required information? What channels of communication can the actor use to communicate with the actor?
3. Imposing Consequences	**Which factors may restrict the ability of the forum to impose consequences?** What information does the forum require to comprehend and assess the justification? What options does the forum have to impose consequences? What possibilities for change do these consequences provide?
4. Effecting Change	**Which factors may restrict the ability of the actor / forum to effect change?** What influence can the forum exercise on the algorithmic system? What other stakeholders are required to effect change? What technical, social and procedural limitations may shape the changes required?

and clarify the cases where the limited agency of forum and actor definitively prohibits such a successful process.

The A^3 framework does not have a legal basis for its application, but rather represents a tool for organisations and outside stakeholders to voluntarily evaluate existing and future algorithmic accountability processes. As such, it is meant to be used as a qualitative and explorative tool in conjunction with other, quantitative assessment tools (see the following Section 6.1 for some examples). To this end, the questions are formulated in a deliberately general and overarching way; they are meant to facilitate a critical discourse on the four phases of the accountability process rather than explicitly address specific deficits of a given context. This allows for a much-needed collaboration between different stakeholders of the system, from domain experts like system developers or management staff, to advocacy groups, academics analysing existing algorithmic systems or policy experts inside and outside of the organisation employing the system. Tailoring the guiding questions to these different groups should be done in a second step after a preliminary, collaborative evaluation of a given system has been done, to facilitate the development of concrete changes that benefit the accountability process. Separating this step from the preliminary discourse is necessary to preserve the wide applicability of the framework in the first step; addressing specific actors must be tied to the context of the system and will inevitably limit the applicability to the specific case.

6.1. *Evaluating the framework in context with modern approaches to algorithmic accountability*

As algorithmic accountability gains more and more attention by academic scholars and practitioners alike, a number of frameworks and approaches have been proposed to address the issue. The following sections aims to compare these existing proposals to the A^3 framework and argue for its value as an alternative point of view.

Starting at a conceptual level, two general approaches to (algorithmic) accountability can be discerned: Accountability as a *virtue* and as a *mechanism* (Bovens, 2010). While academic discourse on the former yields important distinctions for the meaning of accountability in different contexts (e.g., Fourcade et al., 2021a; Fourcade et al., 2021b), few concrete suggestions for the operationalization of accountability as a relational process can be gleamed from these contributions. The latter approach - accountability as a *mechanism* - promises more practical guidelines to assessing and implementing accountable algorithmic systems. Brandsma and Schillemans (2013) propose the *Accountability Cube* as a quantitative assessment tool to compare different systems. While certainly usefuls as a purely evaluative tool, the methodological approach of quantifying the three dimensions of *information, discussions* and *consequences* which are also the foundation of the A^3 framework offers no good indication of *why* certain dimensions may be deficient in a given system. The value of the framework, as stated by the authors, thus lies in the facilitation of *"[...] empirically informed normative judgments on the state of accountability of, for instance, networks, public bodies, or international organizations such as the EU."* (Brandsma and Schillemans, 2013, p.2). Similarly, Tagiou et al. (2019) present a slightly more complex assessment framework for algorithmic accountability along the two dimensions of *organisational* and *algorithmic* issues. While their tool allows for a more fine-grained, assessment, it also prioritizes quantitative evaluation over qualitative inquiry, and operates on a more vague definition of accountability.

What unites many of the existing approaches is a macro-view on accountability processes on an organisational policy level. Busuioc (2021), for instance, identifies largely systemic challenges to algorithmic accountability, including *information asymmetries*, the *inherent opaqueness* of deep learning techniques and the *behavioural effects* of algorithm outputs on human decision-making. Similarly, Buhmann et al. (2020) address *reputational concerns, engagement strategies* and *discourse principles* through four principles (*participation, comprehension, multivocality*, and *responsiveness*). Both of these contributions serve well as examples of organisation-level frameworks to assess and conceptualize algorithmic accountability challenges and solutions, and do offer some more concrete suggestions for intervention, but do so from the larger-scale view of organisational and management perspectives, limiting the applicability of their findings for concrete, human-centered

relational accountability processes as outlined in the two case studies in the previous sections.

Finally, a number of contributions to the literature consider algorithmic accountability from a legal standpoint, often specific to the national legislative context. To name but one example, Engstrom and Ho (2020) take a closer look on algorithmic accountability in the administrative state and suggest oversight boards as a concrete measure to improve the (legal) situation in the USA. While contributions from a legal perspective are particularly relevant for algorithmic technology employed by governmental agencies, the logics of governmental bureaucracy are not always applicable to other contexts (such as private business), and - as argued in the introduction - don't necessarily help organisations implement concrete measures supporting micro-accountability.

Based on these observations, the proposed A^3 framework offers the following advantages. First, agency as a lens offers a wide applicability, from a macro-level of inter- and intra-organisational policy (when conceptualising forum and actor as organisational entities) to the micro-level of human-to-human accountability processes. Secondly, the focus on qualitative evaluation encourages not only the identification of deficiencies in agency, but also offers a chance at insights to develop specific measures providing said agency to actor and forum alike. The open nature of the guiding questions prioritizes critical discourse over quantitative assessment, and provides a structure to interrogate current and future accountability practice for algorithmic systems along the four phases of the accountability process. As such, the A^3 framework should not be seen as a replacement, but rather as a complementary tool to other, more policy-oriented and quantitative frameworks.

7. Limitations and outlook

Accountability is a complex issue that defies simple and straightforward solutions; the agency-based analytic lens should not be misunderstood as the framing device that will 'solve accountability'. Comparative case studies as described by Knight (2001) do not offer definitive results, but represent heuristic approaches and can not *"[...] prove the validity of generalizations beyond providing persuasive evidence in particular contexts."*(Knight, 2001, p. 2). The fact that the two systems being compared - given their different contexts, stakeholders, and technologies - still lend themselves as illustrative examples to show the importance and relevance of agency in the accountability process nevertheless suggests the potential of the approach. Further research to apply and refine the model to other algorithmic systems and contexts is surely needed to contribute to the A^3 framework, and may yet uncover further issues relating to accountability and agency. In particular, comparing agency-related observations in systems that are vastly different in their technologies, but relate to the same context - for instance, predictive policing and facial recognition for law enforcement - may show particular promise due to their overlap in stakeholders.

As a particular point of relevance, the question of legal accountability and the applicability of the A^3 framework to this context deserves attention. While accountability in a legal sense is the prevalent interpretation of the term across many disciplines (including, but not limited to, legal studies itself), the A^3 framework does not specifically cater to this definition. Nonetheless, on a more conceptual level, a reasonable argument can be made that the legislative branch of a (democratic) government in and of itself aims to improve the agency of the executive branch by creating legislature that allows regulation of algorithmic systems and prosecution of offenders against the rule of law. Through this lens, the framework's guiding questions, while in need of adaption in their specific wording, are still applicable. To give but one example: when considering the state's agency to request information, the question *"How can the forum be made aware of the existence of the system?"* may reveal the complexity of identifying private businesses deploying high-risk automated decision making systems, and may point in the direction of specific legislation mandating the disclosure of the development and use of such systems to the government. An example for such a regulation would be the Canadian governments Directive on Automated Decision Making (Canada, 2019), which requires such a disclosure for government agencies. A substantive analysis and adaptation of the framework to the legal context is certainly a desirable avenue of inquiry for further studies, but transcends the scope of this publication.

Lastly, the accountability process as outlined in this publication does not regard the algorithmic system itself as an actant in the sense of Actor-Network-Theory (ANT, Michael, 2016). A re-conceptualization of the accountability process, including the system itself as an active participant in the process, represents a logical next step that takes into account the growing sophistication of algorithmic interfaces and the growing number of artificial agents. A good starting point for such an analysis would be the work by Floridi and Sanders (2004) on the morality of artificial agents, wherein they propose to separate the concepts of moral action and moral responsibility for artificial agents (for instance, ADM systems).

8. Conclusion

Algorithmic accountability remains a wicked problem that presents as a multi-faceted, context-sensitive challenge. Along with issues of transparency, algorithmic literacy, and trust, the accountability of automated systems will remain an important topic for future research. This paper contributes to the discussion by proposing the use of an analytic lens to analyse algorithmic accountability measures and their suitability: *actor* and *forum agency* as part of the accountability process described by Bovens (2007). Two case studies of different algorithmic systems - the Austrian Public Employment Service's AMAS system (Allhutter et al., 2020a) and the EnerCoach energy accounting tool (Cech, 2021) serve as illustrative examples of these accountability processes, and underscore the importance of actor and forum agency to hold algorithmic systems and their results to account.

A critical analysis of the AMAS system reveals the fragility of the accountability process with a diminishing agency of actor and forum alike as it progresses, and shows the inadequacy of the measures aimed at explainability (in the form of explanatory texts) to provide meaningful accountability. The detailed exploration of a hypothetical accountability process between jobseeker and AMS worker reveals the numerous obstacles and high requirements of algorithmic literacy and domain knowledge required to allow a successful conclusion - one that would end in (at least) the potential for change.

In comparison, the EnerCoach case study shows the value of choosing participatory design methodologies that themselves empower and provide agency to the participants, leading to more meaningful and well-tailored measures. The illustrative accountability processes detailed in Section 4 shows how participatory design allowed stakeholders to identify the most problematic aspects, and co-design solutions directed at benefiting the agency of both the *actor* and the *forum* in holding the system to account: visualizations as a reference for expert users, and annotations to help users trace their inputs clearly to the outputs of the system.

Based on the comparison of these two case studies, the A^3 framework offers a set of guiding questions for each of the four steps in the accountability process: *Requesting Information, Providing Account, Imposing Consequences* and *Effecting Change*. While not a specific accountability measure in and of itself, the framework has broad applicability to a wide range of algorithmic systems and the accountability processes required within them, and can help practitioners to approach a satisficing solution to a wicked problem in individual cases. Together with the methodological approach to designing accountability measures, the A^3 framework can help fill the gap between system analysis and developing social, technical and procedural accountability measures, with the ultimate goal to empower the human actors working with and forums affected by an accountable algorithmic system.

The *wicked* nature of the issue of algorithmic accountability defies

simple solutions. Addressing this challenge requires new methodological approaches, including participatory methods and qualitative inquiry, as well as the willingness of different system stakeholders to engage in a critical discourse on the topic. The A^3 framework represents an important step in that direction, by facilitating such an exchange across disciplinary and organisational boundaries. The approach presented in this study can not guarantee the willingness of involved stakeholders to engage in such a process the same way regulation or legal requirements can. Nonetheless, the fact that many organisations already recognize the necessity for increased algorithmic accountability, but lack the methodological framework to approach the issue, indicates the potential for the A^3 framework to fill this gap and empower organisations to improve the accountability of the systems they employ.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Allhutter, D., Cech, F., Fischer, F., Grill, G., & Mager, A. (2020a). Algorithmic profiling of job seekers in Austria: How austerity politics are made effective. *Frontiers in Big Data*, 1–28. https://doi.org/10.3389/fdata.2020.00005

Allhutter, D., Cech, F., Fischer, F., Grill, G., & Mager, A. (2020b). Der AMS-Algorithmus: Eine Soziotechnische Analyse des Arbeitsmarktchancen-Assistenz-Systems (AMAS),. doi. 10.1553/ita-pb-2020-02.

Ananny, M., & Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*. https://doi.org/10.1177/1461444816676645

Binns, R. (2017). Algorithmic accountability and public reason. *Philosophy & Technology, 31*(4), 1–14. https://doi.org/10.1007/s13347-017-0263-5

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal, 13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Bovens, M. (2010). Two concepts of accountability: Accountability as a virtue and as a mechanism. *West European Politics, 33*(5), 946–967. https://doi.org/10.1080/01402382.2010.486119

Brandsma, G. J., & Schillemans, T. (2013). The accountability cube: Measuring accountability. *Journal of Public Administration Research and Theory, 23*(4), 953–975. https://doi.org/10.1093/jopart/mus034

Brennan, T., Dieterich, W., & Ehret, B. (2008). Evaluating the predictive validity of the compas risk and needs assessment system. *Criminal Justice and Behavior, 36*(1), 21–40. https://doi.org/10.1177/0093854808326545

Buhmann, A., Paßmann, J., & Fieseler, C. (2020). Managing algorithmic accountability: Balancing reputational concerns, engagement strategies, and the potential of rational discourse. *Journal of Business Ethics, 163*(2), 265–280. https://doi.org/10.1007/s10551-019-04226-4

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler, & C. Wilson (Eds.), *Proceedings of Machine Learning Research: vol. 81. Proceedings of the 1st conference on fairness, accountability and transparency* (pp. 77–91). New York, NY, USA: PMLR. http://proceedings.mlr.press/v81/buolamwini18a.html

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *3*(1). https://doi.org/10.1177/2053951715622512205395171562251 – 12

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review, 81*(5), 825–836. https://doi.org/10.1111/puar.13293

Canada, T. B. o. (2019). Directive on automated decision-making. https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

Cech, F. (2021). Tackling Algorithmic Transparency in Communal Energy Accounting through Participatory Design, . In *Proceedings of the 10th International Conference on Communities \and Technologies - Wicked Problems in the Age of TechACM Communities and Technologies Conference* (pp. 258–268). https://doi.org/10.1145/3461564.3461577, 9781450390569.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs, 85*(1), 57–80. https://doi.org/10.1080/03637751.2017.1375130

Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology, 103*(4), 962–1023. https://doi.org/10.1086/231294

Engstrom, D. F., & Ho, D. E. (2020). Algorithmic accountability in the administrative state. *Yale Journal on Regulation, 37*(800).

European Commission, T. (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/?qid=1623335154975&uri=CELEX3A52021PC0206.

Fink, K. (2017). Opening the governments black boxes: Freedom of information and algorithmic accountability. *Information, Communication & Society, 21*(10), 1–19. https://doi.org/10.1080/1369118x.2017.1330418

Fink, M. (2021). The EU artificial intelligence act and access to justice. EU Law Live,. https://eulawlive.com/op-ed-the-eu-artificial-intelligence-act-and-access-to-justice-by-melanie-fink/.

Fitzpatrick, G. (2003). *The locales framework, understanding and designing for wicked problems*. https://doi.org/10.1007/978-94-017-0363-5

Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines, 14*(3), 349–379. https://doi.org/10.1023/b:mind.0000035461.63578.9d

Fourcade, M., Kuipers, B., Lazar, S., Mulligan, D., Henriksen, A., Enni, S., & Bechmann, A. (2021a). Situated accountability: Ethical principles, certification standards, and explanation methods in applied AI. *Proceedings of the 2021 AAAI/ACM conference on AI, ethics, and society* (pp. 574–585). https://doi.org/10.1145/3461702.3462564

Fourcade, M., Kuipers, B., Lazar, S., Mulligan, D., Loi, M., & Spielkamp, M. (2021b). Towards accountability in the use of artificial intelligence for public administrations. *Proceedings of the 2021 AAAI/ACM conference on AI, ethics, and society* (pp. 757–766). https://doi.org/10.1145/3461702.3462631

Froehlich, J., Findlater, L., & Landay, J. (2010). The design of eco-feedback technology. In *ACM*. ACM. https://doi.org/10.1145/1753326.1753629

Giddens, A. (1986). *The constitution of society*. Polity.

Greenhalgh, T., & Russell, J. (2009). Evidence-based policymaking: A critique. *Perspectives in Biology and Medicine, 52*(2), 304–318. https://doi.org/10.1353/pbm.0.0085

Hargittai, E., Gruber, J., Djukaric, T., Fuchs, J., & Brombach, L. (2020). Black box measures? How to study peoples algorithm skills. *Information, Communication & Society, 23*(5), 1–12. https://doi.org/10.1080/1369118x.2020.1713846

Janssen, M., Matheus, R., Longo, J., & Weerakkody, V. (2017). Transparency-by-design as a foundation for open government. *Transforming Government: People, Process and Policy, 11*(1), 2–8. https://doi.org/10.1108/tg-02-2017-0015

Joyce, S., Neumann, D., Trappmann, V., & Umney, C. (2020). A global struggle: Worker protest in the platform economy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3540104

Kemper, J., & Kolkman, D. (2018). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society, 0*(0), 1–16. https://doi.org/10.1080/1369118x.2018.1477967

Kensing, F., & Greenbaum, J. (2012). Heritage - having a say. In *Routledge International Handbook of Participatory Design* (pp. 41–56). Routledge. https://doi.org/10.4324/9780203108543-9

Knight, C. (2001). Human-environment relationship: Comparative case studies. *Environment and Ecology: Methods and Measures: Logic of Inquiry and Research Design*, 7039–7045. https://doi.org/10.1016/b0-08-043076-7/04195-4

Latour, B. (1990). Technology is society made durable. *The Sociological Review, 38*(1_suppl), 103–131. https://doi.org/10.1111/j.1467-954x.1990.tb03350.x

Latour, B. (2007). Reassembling the social. In *OUP Oxford*. OUP Oxford.

Lebiere, C., Pirolli, P., Thomson, R., Paik, J., Rutledge-Taylor, M., Staszewski, J., & Anderson, J. R. (2013). A functional model of sensemaking in a neurocognitive architecture. *Computational Intelligence and Neuroscience, 2013*(4124). https://doi.org/10.1155/2013/921695921695–29

Lomborg, S., & Kapsch, P. H. (2019). Decoding algorithms. *Media, Culture & Society, 15*. https://doi.org/10.1177/0163443719855301016344371985530–17

Michael, M. (2016). Actor-network theory. In *SAGE*. SAGE.

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. In *FAT* '19* (pp. 279–288). https://doi.org/10.1145/3287560.3287574arXiv:1811.01439

Morozov, E. (2014). *To save everything, click here: the folly of technological solutionism*. New York: PublicAffairs.

Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics, 2*(1), 25–42. https://doi.org/10.1007/bf02639315

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. In *Penguin Books Limited*. Penguin Books Limited.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press. https://doi.org/10.4159/harvard.9780674736061

Penz, O., Sauer, B., Gaitsch, M., Hofbauer, J., & Glinsner, B. (2017). Post-bureaucratic encounters: Affective labour in public employment services. *Critical Social Policy, 37*(4), 540–561. https://doi.org/10.1177/0261018316681286

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. *Proceedings of the 2018 CHI conference* (pp. 1–13). https://doi.org/10.1145/3173574.3173677

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155–169. https://doi.org/10.1007/bf01405730

Roelen, A., & Klompstra, M. (2012). The challenges in defining aviation safety performance indicators. *Proceedings of the PSAM 11 and ESREL, Helsinki, Pp.*.10.1.1.729.1404

Sandvig, C., and, K. H. D., & Welch, M. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Journal of Life Cycle Assessment*.

Skirpan, M., & Gorelick, M. (2017). The Authority of "Fair" in Machine Learning. arXivorg, cs.CY.

Strengers, Y. (2011). Beyond demand management: co-managing energy and water practices with Australian households. *Policy Studies, 32*(1), 35–58. https://doi.org/10.1080/01442872.2010.526413

Tagiou, E., Kanellopoulos, Y., Aridas, C., & Makris, C. (2019). A tool supported framework for the assessment of algorithmic accountability, *00. Proceedings of the*

10th international conference on information, intelligence, systems and applications (IISA) (pp. 1–9). https://doi.org/10.1109/iisa.2019.8900715

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology, 31*(2), 842–861. https://doi.org/10.2139/ssrn.3063289

Wagner, B. (2019). Liable, but not in control? Ensuring meaningful human agency in automated decision-making systems. *Policy & Internet, 11*(1), 104–122. https://doi.org/10.1002/poi3.198

Wagner, B., Lopez, P., Cech, F., Grill, G., & Sekwenz, M.-T. (2020). Der AMS-algorithmus. transparenz, verantwortung und diskriminierung im kontext von digitalem staatlichem Handeln. *Juridikum*, (02/2020), 191–202. https://doi.org/10.33196/juridikum202002019101

Wagner, I. (2017). Critical reflections on participation in design,. (pp. 1–36).

Wieringa, M. (2019). What to account for when accounting for algorithms. *Proceedings of the ACM conference on fairness, accountability and transparency 2020* (pp. 1–18). https://doi.org/10.1145/3351095.3372833

Wood, G., Day, R., Creamer, E., Horst, D.v.d., Hussain, A., Liu, S., Shukla, A., Iweka, O., Gaterell, M., Petridis, P., Adams, N., & Brown, V. (2019). Sensors, sense-making and sensitivities: UK household experiences with a feedback display on energy consumption and indoor environmental conditions. *Energy Research & Social Science, 55*, 93–105. https://doi.org/10.1016/j.erss.2019.04.013