Exploring Police Perspectives on Algorithmic Transparency: A Qualitative Analysis of UK Police Interviews

Miri Zilka
University of Cambridge
Cambridge, UK
mz477@cam.ac.uk

Carolyn Ashurst
The Alan Turing Institute
London, UK

Luke Chambers
Northumbria University
Newcastle, UK

Ellen P. Goodmann
Rutgers Law School
Camden, NJ, USA

Pamela Ugwudike
University of Southampton
Southampton, UK
The Alan Turing Institute
London, UK

Marion Oswald
Northumbria University
Newcastle, UK
The Alan Turing Institute
London, UK

ABSTRACT

The UK Government's 'Algorithmic Transparency Recording Standard' is intended to provide a standardised way for public bodies and government departments to provide information about how algorithmic tools are being used. To explore the implications of police use of the Standard, we conducted semi-structured interviews with respondents from across UK policing and commercial bodies involved in policing technologies. Our aim was to identify rewards, risks, challenges for the police, and areas where the Standard could be improved. We find that algorithmic transparency is both achievable for policing, and could bring significant rewards. If the Standard became an integral part of an effort to drive reflective practice across the development and deployment of algorithmic technology, it could help police forces to learn from each other, facilitate good policy choices around technology, and decrease wasted costs. However, participants reported notable concerns, including misperception of the dangers of policing technology, and a worry that the Standard will become an administrative burden rather than a benefit for policing or the public. For successful incorporation, we highlight the need to 1) clearly define what is covered by the Standard, 2) provide suitable exemptions for sensitive contexts and tradecraft, 3) ensure that forces have the resources and ability to comply with the Standard, and 4) address supplier responsibilities regarding transparency in procurement contracts. We suggest that future evaluation of the Standard is needed to investigate: a) whether the Transparency Reports created using the Standard meet the needs of intended users, including impacted individuals, advocacy groups, researchers, and legal and policy advisers, b) whether the Standard contributes to an improvement in the quality of policing technology, and c) whether the Standard enables the assessment of the lawfulness of technology used by the police.

EAAMO '23, October 30–November 01, 2023, Boston, MA, USA

ACM ISBN 979-8-4007-0381-2/23/10.
https://doi.org/10.1145/3617694.3623246

ACM Reference Format:
Miri Zilka, Carolyn Ashurst, Luke Chambers, Ellen P. Goodmann, Pamela Ugwudike, and Marion Oswald. 2023. Exploring Police Perspectives on Algorithmic Transparency: A Qualitative Analysis of UK Police Interviews. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23), October 30–November 01, 2023, Boston, MA, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3617694.3623246

1 INTRODUCTION

Data-driven technologies are increasingly informing decision making in justice systems including those in the UK, Australia, Canada, and the United States [11, 12, 19, 35, 36, 57]. Algorithmic tools, including predictive data-driven algorithms, are deployed by police forces to assist in both proactive and reactive crime control [10, 29, 41, 47]. Although these algorithms promise potential benefits for law enforcement—including support for improved decision-making, efficiency, and consistency—there is also a body of work demonstrating the risks posed by algorithmic bias [3, 9, 17, 47]. This includes a lack of contextual understanding or common-sense reasoning, the use of data that encodes inequalities [14, 28, 46], harmful feedback loops [14, 28], and lack of *transparency* [55].

Transparency can improve trust in the motives of the police and enhances the legitimacy and normative acceptance of police directives [50]. It is particularly important in the UK due to the *policing by consent* principle, which means that UK police legitimacy stems from public consent, as opposed to the power of the state. This principle is a long-standing philosophy of British policing [37], and is still commonly cited when a change in legislation alters the powers and responsibilities of the police [1]. Studies of public perception of police legitimacy in the UK [16], and internationally [50], show that transparency (including explaining policing decisions to demonstrate trustworthy motives) enhances perceived police legitimacy, and acceptance of decisions as procedurally fair even if they are unfavourable. For instance, a recent UK study found that trust and legitimacy are essential for public acceptance of live facial recognition technologies used by the police [5].

While academics, charity-commissioned reports, and freedom of information requests have identified some of the algorithmic tools used in UK policing [26, 27], there remains significant opacity around which tools are still in use, and why and how they are deployed [57]. According to the House of Lords Justice and Home

Affairs committee inquiry report: 'There are no systematic obligations on individual departments, public bodies, and police forces to disclose information on their use of advanced technological solutions (Paragraph 98)' [23].

In an attempt to create a mechanism for improved transparency, the UK Government has recently developed and piloted an *Algorithmic Transparency Recording Standard*. It is designed to facilitate a standardised way of collecting and setting out information about algorithmic tools. At present, the process is voluntary, and completed reports are published in a central gov.uk repository, though teams are also encouraged to publish on their organisation's websites. The adoption of the Standard by police in the UK could greatly improve transparency around the use of new and emerging technology in policing [23]. This study—conducted alongside piloting of the Standard—explores police views of the initial version of the Standard, including the benefits and drawbacks of using the Standard to improve transparency around police technologies. We find that although there is a high-level agreement on the importance of algorithmic transparency for police, there was generally no consensus on how this should be achieved in practice.

Following background (§2) and methodology (§3), we outline the main findings from our interviews (§4), and the conclusions and recommendations we drew from them (§5.1). These results and recommendations have been shared with those developing the Standard; we describe the updates made in response in §5.2. We conclude by suggesting next steps (§5.3), and drawing conclusions for transparency mechanisms more broadly (§5.4).

2 BACKGROUND

2.1 Policing and the use of algorithms in the UK

UK regional police forces have been developing and deploying algorithmic tools for over a decade [57]. The incorporation of new data-driven technologies seems to have been driven by experimentation and a local drive for innovation, backed and sometimes funded by government departments [23, 57]. However, more often than not, these technological advancements have been abandoned shortly after they were put to the test, usually with no formal reason [23].

Earlier efforts focused on purchasing 'off-the-shelf' software. A notable example is PredPol (now known as Geolitica) [43], a system designed to predict where and when crime will occur, which was adopted by several UK police forces [31, 44]. Due to a combination of unsuccessful trials and public controversies [47], the UK police have distanced themselves from this tool [30]. More recently there has been a shift away from purchasing external tools, towards development of custom solutions in collaboration with academia and consultancies, or using in-house resources [57].

The challenges that the UK police have been trying to address through technological solutions are varied. In addition to tools for the prediction of demand for police resources, several forces incorporated individual risk and crime solvability assessments to their pipeline. One example of the former is Durham Constabulary's Harm Assessment Risk Tool (HART), which is used to assess the risk of re-offending. Following an arrest, this tool is used to assist decisions as to whether the individual should be charged or directed to a rehabilitation program [41]. Another example is the Evidence-Based Investigative Tool (EBIT), which aims to predict whether or not a case will be solved given the existing evidence. This is then used to decide which cases are worth investigative resources.

To date, there have been considerable differences between police forces with respect to the extent to which they are transparent about their use of algorithmic tools, and in the mechanisms through which they communicate with the public. Transparency mechanisms have included press releases about incorporation of new technologies, publication of academic articles on the tools [29, 41], dedicated webpages [42], and making ethics committee minutes publicly available [40]. Forces also vary on how forthcoming they are when responding to Freedom of Information (FOI) requests regarding their use of technology [26]. These voluntary mechanisms depend solely on the forces' willingness to be forthcoming, and suffer from a lack of consistency, standardisation and oversight.

2.2 Algorithmic transparency in government

Efforts are being made to render public sector use of algorithmic tools more transparent in various jurisdictions globally [25]. Achieving adequate algorithmic transparency in practice, however, is non-trivial and requires much consideration of the mechanisms and level of detail of disclosure. Several authorities, including in Canada [48] and New York City [32] have mandated levels of algorithmic transparency for government bodies. However, the extent of the disclosure varies significantly. While NYC's directory is relatively bare bones (providing the name of the agency, the tool's name and usage data, and narrative descriptions about the tool's purpose and how it functions), Canada has mandated disclosure of the source code itself for government-owned AI. A number of cities, including Helsinki, Amsterdam [18], and Ontario [38] have established 'AI registries', aimed at documenting 'the decisions and assumptions that were made in the process of developing, implementing, managing and ultimately dismantling an algorithm' [22]. Following Executive Order 13960 on 'Promoting Use of Trustworthy AI in Federal Government' [15], US federal agencies must conduct an annual inventory of their AI use cases, and make these publicly available. Inventories must include information such as a summary of the use case, which AI techniques were used, and where the training data originates. In the UK, no such mandatory requirement exists. Instead, the Government has launched a voluntary process called the *Algorithmic Transparency Recording Standard*.

2.3 The UK Algorithmic Transparency Recording Standard

The UK Government's 'Algorithmic Transparency Recording Standard' (ATRS or 'the Standard') was launched in November 2021 by the Cabinet Office's Central Digital and Data Office (CDDO) [21]. Responding to calls for greater transparency in the use of algorithms and data-driven technologies, the Government announced that the Standard will 'promote trustworthy innovation by providing better visibility of the use of algorithms across the public sector, and enabling unintended consequences to be mitigated early on' [21]. The Standard forms part of the UK Government's National Data and National AI Strategies, in particular Pillar 3 of the AI Strategy, namely 'Governing AI Effectively' [20].

The Standard is aimed at any public authority that uses algorithmic tools in its decision-making process. At the time of writing,

the Algorithmic Transparency Recording Standard 'Hub' [52] consists of (i) the Algorithmic Transparency Recording Standard [51], (ii) guidance on using and completing the Standard [54], and (iii) a collection of published Transparency Reports [53]. To complete the Standard, public sector organisations are asked to fill in a template, which comprises two main sections: 'Tier 1', which includes a short non-technical description of the tool and its use, and 'Tier 2', which includes more detailed technical information.

After the release of the initial (draft) version of the Standard, the Cabinet Office, together with the Centre for Data Ethics and Innovation (CDEI)—a UK government expert body, formerly part of the Department for Digital, Culture, Media and Sport and recently incorporated into the new Department for Science, Innovation and Technology—undertook a 'piloting' process. This consisted of working with public sector organisations willing to complete the template with respect to an algorithmic tool in development or use, and collating feedback on the process.

The first two Transparency Reports from the piloting process were published on 1 June 2022, alongside initial conclusions from the feedback [13]. In July and October four more pilots were run, including two within policing: Hampshire and Thames Valley Police's use of Domestic Abuse Risk Assessment Tool (DARAT), and West Midlands Police's use of a tool for exploratory analysis of sexual convictions. Running alongside these pilots, our study further investigated the implications of the Standard for policing. While the pilots focused on completing the Standard for a specific tool, our study asked participants from policing to take a broader lens, and consider the possible implications of adoption in policing more generally from their perspective.

In order to make the Standard and surround process as effective as possible, it is critical to understand the needs and views of both the *users* of the Standard and the potential *audience* of the resulting reports. In addition to gathering the views of users, the government pilot process also sought input from other stakeholders, such as NGOs, and representatives of the public, and ran an open call for further comment. Within the policing context, we see our work to understand the views of those in policing as a critical first step – for example to understand how the police are likely to use the Standard, and what can be done to best ensure they are completed to a high standard. It is also vital to understand the views of the potential audience; we will be undertaking such work in a separate project due to commence in the coming months. See §5.3 for further details.

3 METHODOLOGY

We conducted semi-structured interviews with respondents from across UK policing and commercial bodies involved in policing technologies. Sixteen respondents participated in research interviews for the project: ten representatives of UK police organisations, and six representatives of commercial organisations, including large consulting firms and independent consultants. Participants were approached based on their knowledge and experience of developing, using or managing algorithmic tools in the police, using a purposive, selective sampling strategy.

Interviews were conducted online during February and March 2022. A qualitative and participatory research approach allowed us to co-identify with police practitioners the operational and policy issues that arise in connection with being transparent about their use of algorithmic tools. We used a semi-structured format for the interviews in order to ensure a broadly consistent line of questioning while permitting some flexibility for the research team to probe particular responses within the respondent's areas of expertise. Appendix A provides the questions we posed. An information sheet and guideline questions were sent in advance so that respondents could understand the purpose of the project and give their informed consent. The ethical committee at the Northumbria University approved this study.

In order for our findings to feed into the next iteration of the Standard, we were bound by the timescales of the piloting process. Therefore, the study is exploratory only as we cannot say that data saturation was reached. However, we observed that latter interviews tended to raise similar themes to earlier interviews. We used a general inductive approach [49] to analyse interview data supported by the use of NVivo, with the aim of deriving theory from data, and also identifying inconsistencies and disagreements between participant responses. We identified recurring themes by way of a preliminary coding process, followed by a more granular analysis to explore particular issues and patterns in further detail.

3.1 Limitations

Before outlining the main findings from the interviews, we would like to highlight the following limitations of our study. While we consulted with a number of forces and organisations, the overall range of forces was limited and therefore our findings may not generalise to all law enforcement agencies, such as specialist units. Furthermore, due to the voluntary nature of participation, respondents may have been more favourable towards transparency in the policing domain than average, and therefore our findings may not represent the full range of views on the Standard within policing. Our respondents also do not span the full range of seniority, and may not (for example) reflect the views of Chief Officers. We also note that not all interviewees had gone through the process of completing the Standard.

Critically, our work seeks to understand the views of the potential *authors* of transparency reports in policing. It is vital these are viewed alongside those of the potential *audience* for reports, such as impacted individuals, advocacy groups, researchers and oversight bodies. There are likely many tensions between the views of these stakeholders and those expressed by our interviewees. We therefore stress the importance of further work to ascertain those views (see §5.3).

4 THEMES FROM THE INTERVIEWS

Our research has revealed the following themes, issues and concerns relating to police engagement with the initial version of the Standard. We have grouped these under six main themes:

(1) Scope and use of the Standard (§4.1),
(2) Benefits of police participation in the Standard (§4.2),
(3) Perception risk and data disclosure issues (§4.3),
(4) Innovation and commercial sensitivities (§4.4),
(5) Explainability, ethical scrutiny and evaluation (§4.5),
(6) Resourcing and implementation concerns (§4.6).

Quotes presented below are anonymised, and the following codes indicate the category of research participant: L = police/law enforcement respondent; and C = commercial sector respondent.

4.1 Scope and use of the Standard

At the point of interviews, the draft Standard defined a general scope (i.e., which tools are within the scope of the Standard) and a 'priority scope' (i.e., which tools should be prioritised for completion of the Standard). These definitions have changed since our study; the original definitions to which the answers below refer is in Appendix D.

4.1.1 Which tools are covered by the Standard? There was consensus amongst the interviewees that the scope of the Standard was unclear and not straightforward to interpret, in that it does not specify the range of tools and applications covered (see Appendix B.1.1). Some interviewees found the definition too broad and were unsure if it meant to include non-operational algorithmic tools, used, for example, in HR or maintenance. One commercial interviewee noted that the questions presented in the Standard itself suggest the scope is geared towards machine learning-based tools rather than the wider range of police software (see Appendix B.1.3).

Interviewees were not sure if the Standard applies to all existing tools or only to new tools (see Appendix B.1.2). One police interviewee pointed out potential difficulties and resistance from the police if they were required to fill out the Standard retrospectively (see Appendix B.1.4). Most interviewees leaned towards prioritising tools or applications having direct impact on individuals in the community (see Appendix B.1.5). The need to tighten the scope seemed particularly important if the Standard was going to be mandated. One interviewee raised concerns that the extent and quality of participation may not be consistent across police forces:

> If it was going to be mandated, there would need to be defined exemptions. I think that's where you may end up with disparity from force to force because as we see with the freedom of information exemptions, some forces will apply them more rigorously than others – L4.

4.1.2 At which point of the development process should the Standard be completed? For new applications, interviewees were unsure at which point in the development process the Standard should be completed. One natural point that came up as an option was at later stages of testing, before operational deployment (see Appendix B.1.6). However, some interviewees felt positively towards the idea of using the Standard, or a 'light-touch' version of the Standard, early in the development process in order to gain trust and public backing for new tools and applications early on. One police interviewee responded to this suggestion with the following:

> [T]hat could only be a positive thing and my immediate reaction was if we're considering doing something and there's an opportunity to submit a 'version light', so to speak, of a submission, and to get some early feedback around some of the implications, I think that has the potential to save a lot of hard work – L10.

Others were against putting information into the public domain at an early stage, due to the experimental nature of the development of new applications (see Appendix B.1.7). Interviewees also highlighted that the Standard submission for a given application will likely need to be updated. However, the frequency at which updates would be appropriate is not clear. One interviewee noted that a fixed review period, e.g., annual, may be too slow for some applications yet too quick for others, depending on one of the frequency the tool is used (Appendix B.1.8).

4.1.3 What level of detail is required by the Standard? Although the Standard has structured questions, the interviewees were not clear on the level of detail required when answering the questions. When asked about how much detail they thought was appropriate, interviewees had differing opinions. Some interviewees felt that it would be appropriate to capture as much as possible within the Standard to maximise transparency (Appendix B.1.9). Other interviewees, however, felt that the level of disclosure should be tailored to what is relevant for the general public (Appendix B.1.10).

4.2 Benefits of police participation in the Standard

4.2.1 Building public trust and confidence. Some interviewees stressed that a key benefit of police participation in the Standard is that it provides the opportunity to demonstrate transparency and improve police legitimacy, crucial in England and Wales where 'policing by consent' is the prevailing model, with public trust and confidence the *sine qua non* [33] as acknowledged by this police interviewee:

> We police by consent, we need the trust and confidence of the community in order to exercise our duties and our powers effectively, and in order for us to have that legitimacy, we need to be transparent in everything that we do, not just with the use of algorithms – L10 (see full quote in Appendix B.2.1).

This interviewee also recognised that public input is necessary. Studies of police legitimacy have found that giving people the opportunity to participate in policing decisions that can affect their lives is a key antecedent of procedurally fair treatment and perceived legitimacy [34]. One commercial interviewee argued that transparency was key to addressing public anxieties (Appendix B.2.2). However, another police interviewee wondered about the Standard's cost/benefit ratio and whether compliance with the disclosure requirements of the Standard will in fact generate public and stakeholder acceptance of policing algorithms (Appendix B.2.3).

An additional benefit of police participation in the Standard was highlighted by a police interviewee who stressed that transparency can improve the technical proficiency of policing technologies, which also serves to augment confidence and acceptance (Appendix B.2.4). Another police interviewee similarly emphasised that compliance with the Standard could enable forces to demonstrate proficient implementation of technology-driven policing and this can enhance police legitimacy, particularly in relation to applications of policing technologies (Appendix B.2.5).

The importance of transparency was reinforced by yet another police interviewee who noted that openness is particularly necessary when developing predictive algorithms, as they may pose greater ethical challenges (Appendix B.2.6). Existing research on the disparate outcomes associated with such algorithms suggest that ethnic minorities can be particularly affected [14]. Providing

information about efforts to prevent or address biases can help build trust and improve police relations with affected communities.

4.2.2 Demonstrating legitimacy and openness. An emerging theme seemed to be that openness via transparent algorithm design and implementation is a key benefit of participating in the Standard. A number of interviewees expressed this view, and in one case, it appeared to be partially influenced by experience:

> [I]t is essential for there to be transparency in the police use of algorithms. If I cast my mind back to prior to me starting this role, there are some things that I know now that I would never have considered before my (current) role, which has really highlighted to me how important it is to be transparent in how we're creating the algorithms and what we're ultimately doing with them – L10.

One commercial sector interviewee, while proposing that transparency be limited particularly in relation to data practices, thought that even a limited approach would yield dividends if the information provided by police forces evolves into a repository of data-driven models and facilitates more efficient development of data-driven tools and thus potentially 'deduplication across the different portfolios' (C6). The interviewee was nevertheless of the view that disclosure should be carefully targeted according to the requirements of specific stakeholders (Appendices B.2.7 and B.2.8).

One interviewee pointed out that another benefit of participating in the Standard is that it creates the opportunity for developing a repository of information on policing algorithms, so that forces can learn from each other (Appendix B.2.9). Other interviewees thought that the Standard can drive more reflective practice within forces, as they consider potential downsides of algorithmic deployment (Appendix B.2.10). For police forces to realise the potential of this kind of reflective practice, one interviewee thought that the Standard alone was insufficient, and more oversight would be needed (Appendix B.2.11). The reflective data and algorithmic practices described here are increasingly encouraged by researchers and others who highlight the merits of ex ante and ex post algorithm audits [6, 45].

4.2.3 Privacy rights and human-rights infringements. In discussing the benefits of participation, alongside the potential to build public trust and legitimacy, respondents also considered the opportunity to reflect on privacy rights. For instance, one interviewee felt that participation in the Standard can encourage police forces to consider current and future privacy concerns which will intensify as key aspects of human life become increasingly digitised (Appendix B.2.12). The same interviewee was of the view that addressing public apprehension towards surveillance should be the starting point for dealing with privacy issues and expectations:

> I think that one of the first hurdles to get over is an increasing fear of surveillance in society – L7.

This underlines the potential adverse impact of discriminatory and unwarranted surveillance associated with certain policing algorithms such as live facial recognition technologies [7] and predictive models [46]; both pose human rights implications and can trigger legal action as well as negative publicity, risking the undermining of public trust and legitimacy.

4.3 Perception risk and data disclosure issues

While a considerable majority of interviewees from both the police and the commercial sector could perceive some benefit from engagement with the Standard, a number raised reputational and operational considerations. We have categorised these considerations as *perception risk* and *disclosure issues*.

4.3.1 Perception risk. Interviewees highlighted the public good that could result from the considered use of algorithms. For instance, one police interviewee highlighted problems with current practices in relation to domestic abuse, and the potential for an algorithmic approach to improve those practices (Appendix B.3.1). While respondents appreciated a need to focus on the risks and possible adverse effects of algorithms, they also thought it important to consider how data analytics could improve policing and how not exploring algorithmic approaches could produce an unwanted result: 'the victim may be exposed to more harm' (L9).

This factor links to a concern expressed by some interviewees that information provided by the Standard could give an incorrect or misleading impression of the technology, particularly at an early stage, thus unduly heightening public concern, stoking protests, and inhibiting investment in innovation (Appendix B.3.2). Discussing a shortcoming of unlimited disclosure, another commercial interviewee remarked that although it can enhance transparency and legitimacy, excessive disclosure can also provide false assurance to the public and glance over potentially harmful impacts (Appendix B.3.3). This same interviewee highlights the important issue of meaningful disclosure or legible explanation required for adequate public understanding of the quality and impact of an algorithm. A police interviewee highlighted the importance of anticipating and preparing for questions about new technology deployments: 'if you release information, the natural question is 'Is it working? What are the benefits? How is it being used?', and at the point that we release the information, we have to be able to answer those questions' (L10).

This issue links directly to the discussion about the scope and content of the Standard, and the question of the point in an algorithm's development cycle at which the Standard should apply (and whether a *'Standard-lite'* for early stages could be beneficial). One commercial interviewee was of the view that public transparency during a project's developmental stages could be problematic, and that being transparent at the early stages of development may be counterproductive (Appendix B.3.4).

4.3.2 Data disclosure issues. Understandably, given the operational space in which the police operate, interviewees discussed sensitive policing contexts and capabilities, for instance counter-terrorism and covert policing. Concerns focused upon safeguarding and effectiveness, rather than on public perception, and on how disclosure may impact individuals' safety (Appendix B.3.5). It was suggested that providing the level of detail required by the Standard might allow certain algorithms to be 'gamed', such as algorithms that focus on solvability of crimes due to their simplicity and small number of input factors: 'If you knew all of that [what the algorithm does], you could commit a crime, being very careful around very few pieces of evidence' (C5). It was argued however that 'gaming' was not likely to be a significant issue in other contexts: 'in domestic abuse I don't

see it being gamed anyway, not effectively because emotion is too high often in these cases' (C5).

Furthermore, any requirement to disclose all algorithms publicly was said to risk revealing capability that the police do not have (L7), thus breaching the operational principle of *neither confirm nor deny* applied to sensitive operational techniques. A non-public version of the Standard, linked to other methods to ensure legitimacy such as independent oversight, was suggested as a method of tackling these concerns (Appendix B.3.6). Other interviewees supported a tiered approach, allowing for only partial disclosure with the public when needed (Appendix B.3.7).

4.4 Innovation and commercial sensitivities

4.4.1 Risk of discouraging technology development. Interviewees identified a potential trade-off between implementing the Standard and adopting useful new tools because the rigours of the Standard would raise barriers to innovation. Of note are two kinds of barriers. One is the burden the Standard process places on public officials (Appendix B.4.1). A police interviewee thought that the stringent review process is 'more likely to lead to responsible use of data' but at the same time, might reduce recourse to 'tried and tested solutions to safeguard vulnerable people' (L5). Of course, this comment assumes the tools under review are actually tried and tested, which may not be the case in practice. The second kind of burden the Standard process might place on innovation involves private sector partners, due to potentially high costs of complying with tight regulation (Appendix B.4.2).

4.4.2 Supplier responsibilities. Among interviewees who are working with a third-party supplier to develop and implement algorithmic tools, it was apparent that the responsibilities for explaining and otherwise making the tools transparent reside jointly in the supplier and the public sector customer. The balance of responsibility may depend on how much involvement the force has in developing the tool on the back end, and then how much the force tailors the tool or its implementation on the front end (Appendix B.4.3). One police interviewee said that as technology companies can create high-complexity tools, they would expect the supplier to provide a transparent on-boarding, allowing the forces to understand what 'they're getting into' (Appendix B.4.4).

It was seen as important to define up front how to allocate responsibility for assessing and explaining model design and performance over time (Appendix B.4.5). One police interviewee suggested that police forces get help from the Police Digital Service in clarifying public and private sector responsibilities for transparency-related tasks (Appendix B.4.6). Allocating responsibility for transparency to private sector suppliers raises another issue, discussed below, which is the possible resistance to disclosure those suppliers might exhibit due to business concerns.

4.4.3 Trade secrets and commercial sensitivity. Interviewees identified private sector supplier resistance to transparency as a potential obstacle to implementation of the Standard. In particular, they highlighted concerns with proprietary trade secrets. In the absence of intellectual property protection, which algorithms are unlikely to have in the UK, there is 'nothing to prevent once you publish the actual algorithm itself, somebody else just taking it and using it or selling it . . . Until that changes, which I don't think it will, you wouldn't be able to convince some places to actually publish what they're doing' (C5). The degree to which supplier concerns will impede transparency may depend on the attitudes of those suppliers, their market power, and the contractual arrangements that address this issue. One of the suppliers who participated in the project said that it does not 'develop any software that we won't share with our police and partners'. This interviewee acknowledged that other suppliers are different: 'They do develop code and that is commercially sensitive. That's effectively the unique selling point and they're probably not going to be willing to share that code base' (C6).

4.5 Explainability, ethical scrutiny, and evaluation

4.5.1 Explaining the technology/explainability. Concern was raised as to how best to explain technical capabilities in a succinct way that the general public can readily understand (Appendix B.5.1). One police interviewee highlighted a risk that the Standard could result in information overload for the public:

> I think we've probably provided more details than the public needs at the moment, or can handle – L5.

Explainability of the technology for the police as well as the public was highlighted as a desired outcome, with one interviewee highlighting the potential value for internal users (Appendix B.5.2).

The public may also *overestimate* police abilities or infrastructure, and transparency laying bare true capabilities may impact public trust (Appendix B.5.3). The issue of public trust is reflected in the discussion during the interviews. Though attitudes towards transparency were generally positive, some participants felt that in some cases compliance could *reduce* public trust. Such interviewees were concerned about which information should be shared, and how, in order to avoid causing confusion to the public. The above interviewee believed that the public was likely not to understand the terminology without substantial background being offered. Furthermore, the multiplicity of terms used by different police forces could result in inconsistency in respect of the completion of the Standard. This was expanded on later in the interview, with the interviewee responding positively to a suggestion for a glossary and stating:

> [I]t's essential because otherwise you're just going to get 43 different languages and slight variations or preferences of how people would describe things and I would probably go a step further and have common definitions of things like a nominal or what master data management is – L5.

A commercial interviewee agreed that a 'single considered, consolidated way of communicating' via a glossary and/or methods of communication linked to the Standard would be of benefit (C2).

4.5.2 Ethical scrutiny. Despite a significant agreement that ethical scrutiny is important, interviewees had a range of opinions on the time, range, and scope of such scrutiny.

One commercial sector interviewee was concerned that the burden of ethics processes and anticipation of projects being stopped can stifle innovation (Appendix B.5.4). Although these fears of ethical scrutiny shutting down projects do not reflect the reality of ethical scrutiny processes (such as those operated by West Midlands

Police [40]), it is worth acknowledging that these fears could have a real effect. If forces or commercial providers were to choose not to innovate in certain areas for fear of ethical scrutiny, a self-imposed chilling effect would be an unintended consequence of that scrutiny.

Other interviewees expressed uncertainty, not with the ethical scrutiny process itself, but regarding the point at which to disclose projects for that scrutiny:

> [A]t what point do you present something to the Ethics Committee? Because sometimes you haven't really got that much to talk about, but people are going to want to know "How accurate is it? Because that leads into how much we support you doing it". But we want to talk about developing it with you to see if you even get to support developing it – L9.

Other interviewees mirrored similar sentiments (Appendix B.5.5).

At a high level, interviewees involved in policing demonstrated an interest in tracking bias and accuracy in algorithmic tools. Our interview data suggests that the interest in detecting and mitigating bias is well aligned with the introduction of the Standard. However, the practicality of achieving high accuracy and low bias is not always straightforward in practice (Appendix B.5.6). Questions were also raised about the communication of accuracy levels and the impression that these might give to the public, as '60% doesn't sound very good at all' (L10). (see the full quote in Appendix B.5.7). This reflects comments from other interviewees regarding the public response to the published standard. In this instance, this is grounded in the uncertainty of the public to be able to interpret metrics such as model accuracy. Some interviewees were concerned that the Standard may facilitate the publication of accuracy and other performance metrics without sufficient context to allow for a robust interpretation of those numbers.

4.6 Resourcing and implementation concerns

4.6.1 Resource required to comply with the Standard. Of significant concern to interviewees was the resource burden of complying with the Standard. The size and type of burden imposed depends on the kind of tool that is being used and at what stage of implementation the force would fill out the template. Depending on competing requirements, resource constraints may prevent adoption of a tool or compliance with the Standard (Appendix B.6.1).

One commercial interviewee was less concerned with resource demands, noting that the template codifies burdens that forces should be undertaking in any case: 'Sure it takes time to write down the answers to those questions, and the answers to those considerations, but we should be doing that anyway' (C5).

4.6.2 Risk of increasing FOI requests. While it might be expected that proactive publication of information would result in a reduction of the number of freedom of information (FOI) requests received by policing bodies on the topic of algorithmic tools, interviewees were not convinced by this assumption. One police interviewee told us that they were receiving:

> [A]n increasing number of FOIs coming from a variety of different sources—media and academia asking some quite detailed questions about police approach to AI algorithms, etc. Who's using what, who's trying what? What are you thinking about using? So certainly it's getting noticed and I think that there's a huge risk of giving people sufficient information to get concerned, but not enough to actually satisfy themselves that it's not as bad as they think it may be (L7).

Another policing interviewee however regarded dealing with FOI requests as 'an accepted part of doing business', and was less concerned about the potential resourcing implications (L3).

5 DISCUSSION

5.1 Study conclusions and recommendations

5.1.1 Rewards and risks of the Standard for police forces. Interviewees generally thought that the rewards of a carefully tailored Standard implemented at the right stage of algorithmic development outweighed the risks, provided that the identified challenges were addressed. These *rewards* centre on:

- Opportunities to demonstrate the legitimacy of policing technology, and build public confidence and trust. Interviewees highlighted the public good that could result from the careful use of algorithms.
- Increased public interest in, and understanding of, policing algorithmic tools.
- Increased sharing of best practices, and potential pitfalls, among police forces, and therefore an opportunity to improve policy choices around technology, and decrease costs.
- Increased thoughtfulness among police force personnel about building and implementing new tools, resulting in an improvement in the quality (and societal outcomes) of deployed technology.

The primary *risks* are:

- An increased misperception of the dangers of policing technology if the use of algorithmic tools are not appropriately compared to the status quo and current methods.
- A possible increase in public opposition and thus increased pressure to turn away even from useful innovation.
- A potential ability to 'game' an algorithm in limited contexts; this risk was not seen as particularly significant by all interviewees and only applicable to certain applications.

The primary *challenges* are:

- Ensuring that forces have the resources to comply with the Standard and also to respond to the increased public interest that could ensue.
- Ensuring that supplier responsibilities to assist the police with compliance with the Standard are factored into commercial arrangements.
- Factoring compliance with the Standard into other oversight processes, including independent oversight.
- Ensuring that the Standard allows sufficient scope for explaining the issues around current methods, and the potential for technology to improve the status quo.

5.1.2 Improving the Standard and surrounding process. Based on our analysis of the interviews, we suggested the following key areas for amendment and improvement of the initial version of the

Standard. These recommendations were included in a report we shared and discussed with the developers of the Standard.

- Interviewees agreed the *scope of the Standard* was unclear, and required considerable clarification regarding how to decide whether a tool or application were covered by the Standard. Please find an additional discussion around our interpretation of the scope and its implications in Appendix C.
- To provide clarity, and to deal with the resourcing concerns, a *list of tools in scope* could be produced (initially including the most high-risk tools, such as those that produce individualised risk/predictive scores, and those which inform operational deployments or evidential stages). This would enable the application of the Standard to be ramped up over time, and avoid forces having to interpret terms such as 'complex' processing.
- Many interviewees agreed that the Standard would benefit from a more substantial *glossary* of relevant policing terms, and additional guidance on the *level of detail* required in each section, and how *accuracy rates* should be described, justified, and explained in order to ensure consistency.
- Interviewees expressed different views on the *stage of project development at which the Standard should apply, and at which point(s) submissions should be updated.* This would need to be decided after further discussion with policing bodies. Benefits could arise from transparency at relatively early stages as set out above. We therefore suggest that consideration is given to a *'Standard-Lite'* which could be used for early stage projects.
- Clarity is needed as to whether the Standard will be *mandated* for policing (either now or in the future) and if so, how *exemptions and issues of sensitive tradecraft* will be handled. Our interview data does not suggest any overriding reason why the Standard should not be applied in policing, subject to suitable exemptions and reasonable flexibility due to resourcing pressures. Limiting the scope initially to tools on a defined list could assist in mitigating concerns over sensitive policing contexts and capabilities.
- Consideration could be given to a *non-public version of the Standard*, available for sensitive applications and tools as tightly defined, and available for review to independent oversight bodies in order to ensure legitimacy.
- In order for the Standard to contribute to *improving the quality of policing technology*, it should be linked to *methods of oversight and promotion of best practice on a national basis*, and used to *enable police forces to learn from each other*. Otherwise, the Standard may come to be regarded as an administrative burden rather than a benefit for policing.
- To support police compliance with the Standard, *supplier responsibilities* – including appropriate disclosure of algorithmic functionality, data inputs and performance - should be covered in procurement contracts and addressed up front as a mandatory requirement of doing business with the police.

5.2 Updates to the Standard in response to our findings

In response to the CDDO-CDEI pilot process and the findings from our study, the CDDO have released a new version of the Standard (v2.1). The additions and improvements related to our study include:

- An additional field on model performance.
- Additional fields on data completeness, representativeness, and data cleaning.
- Improved explanations in the 'notes for completion' column throughout the Standard.
- Further guidance on the scope, including which tools are applicable.
- Guidance on when to fill out a Transparency Report. Teams are encouraged to begin completion as early as possible during development, but publication is only expected at the point of piloting or deployment.
- Guidance on exemptions in which a Report should not be published (e.g., security, IP, or gaming concerns).
- Guidance on the expected content and level of detail for each section and field, including FAQs, and illustrations.

Additional improvements, such as those relating to the structure of the Standard, additional guidance on the process for completing and publishing the Report, and further amendments to the fields and their names have also been made in response to the pilot process. Details on all changes to the Standard can be found in [8].

5.3 Suggested next steps

We have listed possible improvements to the Standard based on the views of those who might be responsible for writing a Transparency Report (§5.1), and highlighted those that have already been incorporated into the most recent version of the Standard (§5.2). Other forms of evaluation could also be utilised to inform further improvements, as follows.

First, **work to understand the views of those representing the potential *audience* for Transparency Reports should be undertaken**. The potential audience for Transparency Reports includes: Members of the public, particularly individuals who might have been impacted by a particular system; community and advocacy groups representing communities that may be impacted; researchers at academic institutions or policy think tanks; oversight bodies within government (e.g., the Home Office); those working in policing that may wish to understand what tools are being used (such as senior leaders, or other police forces who may wish to learn from best practice); and legal and policy advisers who wish to ascertain the lawfulness of the use of the tool in the given context.

Interviews could be used to ascertain whether the current form of the Transparency Reports meets the needs of such stakeholders, including whether Reports contain the information they require, whether this information is communicated in a useful and clear manner, and what improvements could be made to the process to ensure the utility of Reports from their perspective. A critical aspect is to evaluate whether the Standard in fact enables the sort of transparency that is required to determine if the use of technology by the police is lawful, i.e. whether the Standard helps to address the questions that must be answered to ascertain whether there is legal justification for use of a tool [4, 39, 56]. Key questions for impacted individuals and communities include whether the Reports facilitate informed consent, and what methods need to be in place to facilitate prompted discussion, appeals and complaint. In an upcoming project we will begin this work, by seeking to understand the views of impacted individuals and communities on such topics.

It is likely that the needs and desires for Reports will differ greatly between authors and audience, and there are many potential value conflicts between those in policing and other stakeholders. It is therefore absolutely critical that the views of both are understood as early as possible, so that tensions can be identified, and careful consideration given as to how to address them in the design of the Standard and the surrounding process. In order to ensure that the tools employed by the police are trustworthy, improving Transparency is merely a first step: transparency mechanisms alone are no guarantee of meaningful engagement or accountability [2, 24]. It is therefore vital that work is done to understand how to best design such mechanisms, and what else is needed, through consultation with those who we would hope to see engage with the outputs.

Second, **pilots and interviews could be used to gain a deeper insight into the use of the Standard in different policing contexts**, depending on the intended use (e.g., public-facing, investigation aid, intelligence collection). In particular, the use of the Standard within sensitive contexts requires further investigation. Further testing could also be used to test some of the proposals outlined in this work, such as a 'Standard-Lite' version for early-stage projects, and additional resources such as additional guidance and glossaries. Further work could investigate alternative or complementary processes and methods to enable the police to learn from each other, with respect to effective and ethical algorithmic tools.

Finally, **continual engagement with police and the intended audience is needed to understand their relationship with the Standard over time**: when it is completed and by whom, how it is used, the benefits and costs incurred, and any remaining risks and challenges to mitigate. In particular, if the standard is only adopted by the already forthcoming forces, it may not fully serve its intended purpose. We must continue to evaluate whether the Standard meets its overarching aims: to promote trustworthy innovation, to enable unintended consequences to be mitigated early on, and to improve the quality of technology within policing.

5.4 General implications for transparency mechanisms

This study has highlighted many of the potential benefits of the Algorithmic Transparency Reporting Standard, for both the public and the police, so long as efforts are made to mitigate the potential risks, and provide adequate incentives and support to police forces. While this work has focused on one particular mechanism, we believe many of our findings apply to the development of transparency mechanisms more generally.

In particular, this study highlights the importance of: (i) *clear guidance*, e.g., related to the scope, level of detail, communication of common aspects (such as performance metrics), and when documentation should be created and updated; (ii) careful consideration of *how such mechanisms should be incentivised* (e.g., should they be mandatory); and (iii) *how should such mechanisms be integrated* into surrounding processes, including how they should be linked to other oversight processes, and where responsibilities lie when systems are procured from external vendors.

This work also demonstrates the value of getting 'buy-in' for transparency mechanisms from system owners. Otherwise the risks, costs and challenges can easily result in owners deciding not to use the process, or—if mandatory—to not complete it to a high standard. In either case this can negatively impact the quality of data/information made available. Identifying, articulating and amplifying the potential benefits to potential users, alongside work to understand and reduce potential barriers can help to get users onboard. Work to ensure such processes meet user requirements, and are proportionate to the level of risk will also positively contribute.

Relatedly, there needs to be careful consideration as to if and when such a process should become mandatory. Prematurely mandating a process that has yet to be sufficiently tested and refined risks frustrating and demotivating users. Once refined, there are still risks associated with mandating the process—if users do not see the value but are still required to complete the process they may attempt to game the requirement, or provide the minimum accepted input. Again, this shows that getting users to engage in earnest is a crucial part of getting good quality information. As an alternative to mandatory requirements, other incentives can be used to encourage adoption, such as promoting forces that participate as responsible and tech-savvy, articulating the benefits gained by previous participants, individual incentives such as contributing to bonuses and promotions, and making Transparency Reports a factor in decisions about deployment and procurement of systems. However, given the pressing need to improve trust and trustworthiness of algorithmic systems (in policing and beyond), it may be reasonable to mandate completing the Standard after some time, at least for high-risk or highly contentious applications, such as facial recognition.

ACKNOWLEDGMENTS

We thank the reviewers for their helpful comments and suggestions. M.Z. acknowledges support from the Leverhulme Trust grant ECF-2021-429. E.G. thanks The Knight Foundation.

REFERENCES

[1] Unmesh Desai AM. 2020. Policing with Consent. https://www.london.gov.uk/sites/default/files/policing_with_consent.pdf
[2] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society* 20, 3 (2018), 973–989.
[3] Solon Barocas and Andrew D Selbst. 2016. Big data's disparate impact. *Calif. L. Rev.* 104 (2016), 671.
[4] Janina Boughey. 2023. Transparency in outsourced automated decision-making systems. *Public Law* (April 2023), 206–214.
[5] Ben Bradford, Julia A Yesberg, Jonathan Jackson, and Paul Dawson. 2020. Live facial recognition: Trust and legitimacy as predictors of public support for police use of new technology. *The British Journal of Criminology* 60, 6 (2020), 1502–1522.
[6] Shea Brown, Jovana Davidovic, and Ali Hasan. 2021. The algorithm audit: Scoring the algorithms that score us. *Big Data & Society* 8, 1 (2021), 2053951720983865.
[7] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.
[8] CDDO. [n. d.]. Change log Algorithmic Transparency Recording Standard v1.1 to v2.1. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1125890/Change_log_ATRS_v1.1_to_v2.1.docx
[9] Alexandra Chouldechova. 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data* 5, 2 (2017), 153–163.
[10] Lina Dencik, Arne Hintz, Joanna Redden, and Harry Warne. 2018. Data scores as governance: Investigating uses of citizen scoring in public services project report. (2018).
[11] Metropolitan Police Department. 2022. Facial Recognition. https://www.met.police.uk/advice/advice-and-information/fr
[12] Metropolitan Police Department. 2022. ShotSpotter Data, Disclaimer and Dictionary. https://mpdc.dc.gov/publication/shotspotter-datadisclaimer-and-dictionary
[13] Amy Dickens and Elena Hess-Rheingans. 2022. Piloting the UK algorithmic transparency standard. https://cdei.blog.gov.uk/2022/06/01/piloting-the-national-

algorithmic-transparency-standard/
[14] Danielle Ensign, Sorelle A Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. 2018. Runaway feedback loops in predictive policing. In *Conference on Fairness, Accountability and Transparency*. PMLR, 160–171.
[15] EO13960. 2020. Promoting the use of trustworthy artificial intelligence in the federal government. https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government.
[16] Kathryn Farrow. 2020. Policing the pandemic in the UK using the principles of procedural justice. *Policing: a journal of policy and practice* 14, 3 (2020), 587–592.
[17] Anthony W Flores, Kristin Bechtel, and Christopher T Lowenkamp. 2016. False positives, false negatives, and false analyses: A rejoinder to machine bias: There's software used across the country to predict future criminals. and it's biased against blacks. *Fed. Probation* 80 (2016), 38.
[18] Luciano Floridi. 2020. Artificial intelligence as a public service: Learning from Amsterdam and Helsinki. *Philosophy & Technology* 33, 4 (2020), 541–546.
[19] New South Wales Government. 2022. Facial Recognition. https://www.police.nsw.gov.au/crime/terrorism/terrorism_categories/facial_recognition
[20] UK Government. [n. d.]. National AI strategy. https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version
[21] UK Government. 2021. UK government publishes pioneering standard for algorithmic transparency. https://www.gov.uk/government/news/uk-government-publishes-pioneering-standard-for-algorithmic-transparency
[22] M Haataja, L van de Fliert, and P Rautio. 2022. Public AI registers: realising AI transparency and civic participation in government use of AI. 2020.
[23] Justice, Home Affairs Committee, et al. 2022. Technology Rules? The advent of new technologies in the justice system. *HLPaper180. Westminster: The House of Lords* (2022).
[24] Jakko Kemper and Daan Kolkman. 2019. Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society* 22, 14 (2019), 2081–2096.
[25] Pascal D König and Georg Wenzelburger. 2021. The legitimacy gap of algorithmic decision-making in the public sector: Why it arises and how to address it. *Technology in Society* 67 (2021), 101688.
[26] Liberty. 2019. REPORT: POLICING BY MACHINE. https://www.libertyhumanrights.org.uk/issue/policing-by-machine/.
[27] Liberty. 2021. New technologies and the application of the law: Written evidence. https://committees.parliament.uk/writtenevidence/38701/pdf/
[28] Kristian Lum and William Isaac. 2016. To predict and serve? *Significance* 13, 5 (2016), 14–19.
[29] Kent McFadzien, Alan Pughsley, Andrew M Featherstone, and John M Phillips. 2020. The Evidence-Based Investigative Tool (EBIT): a Legitimacy-Conscious Statistical Triage Process for High-Volume Crimes. *Cambridge Journal of Evidence-Based Policing* 4, 3 (2020), 218–232.
[30] BBC News. 2018. Kent Police stop using crime predicting software — bbc.co.uk. https://www.bbc.co.uk/news/uk-england-kent-46345717
[31] Patricia Nilsson. 2018. First UK police force to try predictive policing ends contract. https://www.ft.com/content/b34b0b08-ef19-11e8-89c8-d36339d835c0 Accessed: 2022.
[32] NYC. 2022. Algorithms Management and Policy Officer. https://www1.nyc.gov/site/ampo/index
[33] College of Policing. 2022. Professional standards: Authorised Professional Practice. https://www.college.police.uk/app/professionalstandards
[34] College of Policing. 2022. Professional standards: Authorised Professional Practice. https://www.college.police.uk/app/professionalstandards
[35] Office of the Australian Information Commissioner. 2021. AFP ordered to strengthen privacy governance. https://www.oaic.gov.au/updates/news-andmedia/afp-ordered-to-strengthen-privacy-governance
[36] Office of the Privacy Commissioner of Canada. 2021. Police use of Facial Recognition Technology in Canada and the way forward. https://www.met.police.uk/advice/advice-and-information/fr
[37] The Home Office. 2012. Definition of policing by consent — gov.uk. https://www.gov.uk/government/publications/policing-by-consent/definition-of-policing-by-consent
[38] Ontario. 2022. Data Catalogue. https://data.ontario.ca/group/artificial-intelligence-and-algorithms
[39] Marion Oswald. 2018. Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2128 (2018), 20170359.
[40] Marion Oswald. 2021. A three-pillar approach to achieving trustworthy and accountable use of AI and emerging technology in policing in England and Wales: Lessons from the West Midlands data ethics model. *Forthcoming in European Journal of Law and Technology* (2021).
[41] Marion Oswald, Jamie Grace, Sheena Urwin, and Geoffrey C. Barnes. 2018. Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality. *Information & Communications Technology Law* 27, 2 (2018), 223–250. https://doi.org/10.1080/13600834.2018.1458455
[42] Metropolitan Police. 2023. facial recognition | metropolitan police. https://www.met.police.uk/advice/advice-and-information/fr/facial-recognition/
[43] PredPol. [n. d.]. How Predictive Policing Works. https://www.predpol.com/how-predictive-policing-works/
[44] Predpol. [n. d.]. Kent Police Use PredPol To Prevent Violent Crime. https://www.predpol.com/kent-police-use-predpol-to-prevent-violent-crime/ Accessed: 2022.
[45] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 33–44.
[46] Rashida Richardson, Jason M Schultz, and Kate Crawford. 2019. Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *NYUL Rev. Online* 94 (2019), 15.
[47] Aaron Sankin, Dhruv Mehrota, Surya Mattu, and Annie Gilbertson. 2021. Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them. https://themarkup.org/prediction-bias/2021/12/02/crime-prediction-software-promised-to-be-free-of-biases-new-data-shows-it-perpetuates-them
[48] Treasury Board Secretariat. 2020. Directive on automated decision-making. *Ottawa (ON): Government of Canada (modified 2019-02-05* (2020).
[49] David R Thomas. 2003. A general inductive approach for qualitative data analysis. (2003).
[50] Tom Tyler. 2017. Procedural justice and policing: A rush to judgment? *Annual review of law and social science* 13 (2017), 29–53.
[51] UK Government. 2023. Algorithmic Transparency Recording Standard. https://www.gov.uk/government/publications/algorithmic-transparency-template
[52] UK Government. 2023. Algorithmic transparency recording standard hub. https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub
[53] UK Government. 2023. Algorithmic Transparency Reports. https://www.gov.uk/government/collections/algorithmic-transparency-reports
[54] UK Government. 2023. Guidance for organisations using the algorithmic transparency recording standard. https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard
[55] Ali Winston. 2018. Palantir has secretly been using New Orleans to test its predictive policing technology. *The Verge* 27 (2018).
[56] Karen Yeung and Adrian Weller. 2018. How is 'transparency' understood by legal scholars and the machine learning community? (2018).
[57] Miri Zilka, Holli Sargeant, and Adrian Weller. 2022. Transparency, Governance and Regulation of Algorithmic Tools Deployed in the Criminal Justice System: A UK Case Study. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (Oxford, United Kingdom) *(AIES '22)*. Association for Computing Machinery, New York, NY, USA, 880–889. https://doi.org/10.1145/3514094.3534200

A INTERVIEW GUIDE

Introduction

Interviewer to provide a short summary of the project, introduce the team and provide assurance that responses will remain non-attributable.

- Before we begin, do you have any questions about the project?
- What is your role in respect of the development and use of policing algorithms?

Transparency and algorithms

- What is your view of the need for transparency in respect of algorithms used by policing?

Completing the draft standard

- What was your role in completing the draft standard? Who else needed to be involved?
- Do you think that others should be involved?
- Did you understand the definition of algorithmic tools that are covered by the standard?
- Did this definition exclude any algorithms that you use within the force?
- What is your overall view of the process of completing the standard?
- When do you think is the right stage for a tool to be disclosed via the standard? How often should this be updated? What resource implications would this have?
- Which sections were hard to answer and why?
- Did you have concerns about disclosing particular information in answer to any of the questions and why?

Benefits and concerns

- What organisational concerns were expressed about publication of the standard?
- What benefits could accrue from applying the transparency standard?
- What issues or challenges could accrue from applying the transparency standard?
- What strategic, practical and policy issues will need to be considered?

Wrap-up

- Do you have any questions about the interview?
- Interviewer to reaffirm that responses are non-attributable

B INTERVIEWEE QUOTES

B.1 Scope and Use of the Standard

B.1.1 .

Defining the threshold, defining when an Excel macro becomes something that we need to be transparent about, is part of the problem… much like other forces, I did find it difficult to define when this applies – L6.

B.1.2 .

I'm not entirely sure what they're after, whether they're just after new ones, or whether they're also after the myriad of tools that are used by police every day – C5.

B.1.3 .

One of the later questions seemed to assume that you were dealing with a system which learned from data or a system which was making decisions, but some of the things that potentially were in scope, like some of the technology we're looking at, doesn't have a learning model – C2.

B.1.4 .

Do you apply retrospectively or do you draw a line in the sand and move forward? Any new technologies or any new use of algorithms we're going to disclose but we're not going to do this retrospectively – L5.

B.1.5 .

In terms of policing activity, you know project insights which impact on people, which have the ability of or possibility of significantly impacting on an individual or group of individuals has to be part of this scrutiny – L10.

B.1.6 .

'So I think we've got a fairly classic development process anyway between development then testing and before we move into live production. And I suppose this standard would come towards the end of that testing phase. So we want to complete it before we move things into production or into live use because it's something we want to have undertaken before we use something in earnest – L6.

B.1.7 .

we haven't been through an ethical board because we're not going live with it next week. It's not operational yet. It's an experiment form. It's in a controlled environment and so it hasn't even been decided that it will be operational yet. And so I would be a little bit wary about disclosing the development of something before it's actually been decided that it will be operationally used and before it's gone through the internal ethical considerations about whether it's appropriate – C2.

B.1.8 .

(regarding annual updates) in some cases that's going to be too slow and in some cases it's going to be too fast and the length of time has to be related to the tool that you're using, because a tool that's only used for 200 decisions a year, probably doesn't have the data to be reviewed every year, and it's going to be very onerous. But a tool that's going to be used for a million decisions a week, probably needs reviewing more than annually – C5.

B.1.9 .

My starting point would be that we would want to disclose as much as we possibly can because it defeats the point of being transparent if we, from the outset, actively seek to provide the minimum as opposed to providing what we can – L10.

B.1.10 .

There will be certain pieces of information which are only relevant to certain stakeholders – C6.

B.2 Benefits of Police Participation in the Standard

B.2.1 .

[...] We police by consent, we need the trust and confidence of the community in order to exercise our duties and our powers effectively and in order for us to have that legitimacy, we need to be transparent in everything that we do, not just with the use of algorithms. So I think this feeds into that broader need, whether it's the use of force, whether it's stop and search, whatever the activity is within policing. I think it's really important that we from the outset involve individuals, explain, listen to views, and reassure – L10.

B.2.2 .

If we thought that we wouldn't disclose something because a particular individual or group of individuals might be worried that we're using it, I agree that that's absolutely the reason that we should be really onerous around that. And that's something that I would seek to address head on. If people are worried, there's an opportunity there as opposed to a risk, there's an opportunity to be open. There's an opportunity to be transparent. There's an opportunity to explain, to reduce concerns – C6.

B.2.3 .

If this goes ahead, then I do think there should be a review at some point of the level of effort required by forces against the actual take up from the public and interested parties, and I know these things can take a few years before they really gain ground. But, does it actually become a bit of a white elephant? – L2.

B.2.4 .

Every algorithm that's produced with this sort of technology or approach in criminal justice should have a transparency standard attached to it, because that's the only way you start to gain industry standards and confidence – L8.

B.2.5 .

So I think that the main thrust for me of the necessity of this piece of work is around the public competence and legitimacy side of how we blend technology with policing – L7.

B.2.6 .

I don't think it would be useful to have transparency about the fact that police map crime. I would have thought that the majority of people would have expected police to be mapping crime as in crimes that have occurred - if it was about where crimes may occur in the future, that, to my mind, is worthy of going through this transparency process – L7.

B.2.7 .

we can share probably a very low level of detail with the likes of an Ethics Committee, how we're doing all of this analysis and where that information is coming from and how that information was collected – C6.

B.2.8 .

There will be certain pieces of information which are only relevant to certain stakeholders – C6.

B.2.9 .

If there's a public register from our UK policing perspective as a single force, you then get to understand what all the other forces are doing initially, which is quite difficult to grasp at the moment and learn from their successes and failures and in betweens, and I think that's going to be valuable for every force – L5.

B.2.10 .

Even if they're not real negative outcomes, I think identifying potential negative outcomes, which is the thing that we tried to go through with our own process, - we're like "well in the worst case scenario what's the worst possible impact of using this tool?"- and kind of running through some scenarios. So I think that might be a useful part of the standard – L6.

B.2.11 .

'It might also be possible to create a peer reviewing mechanism of sorts. I feel that without this oversight, we will be missing the main benefit of transparency; the ability to ensure that algorithms used in the public sector are up to the task, and are being built properly, with proper attention to the data, and to the drift that occurs with algorithms. Only through this mechanism will organisations and data scientists be forced to remain up to date with the rapid growth in the field of machine learning, especially in the areas of safety, data protection, fairness/bias and machine learning ethics – C5.

B.2.12 .

we have to not only look at our considerations of people's expectations of privacy and security now but moving into the future because I think that our future selves will have an even higher expectation of privacy as more and more of our life starts to occupy a digital space – L7.

B.3 Perception Risk and Data Disclosure Issues

B.3.1 .

…quite a lot of our homicides are domestic abuse related and we're quite bad at seeing the risk quite often because the police officers are dealing with a presenting incident there and then, not the full background, and this is where I think algorithms can help understand the more contextual picture – L9.

B.3.2 .

there's a huge risk of giving people sufficient information to get concerned, but not enough to actually satisfy themselves. It's not as bad as they think it may be… it's really quite a real risk of giving out too much, too soon before we're in a position to defend it and answer it – L7.

B.3.3 .

My main concern stems from my feeling that there are two benefits from transparency. The first is the availability of information to people who the data relates to, and for fueling legitimacy and openness about the kinds of things that algorithms are used for. However, this type of transparency will raise comments based on the perceptions of use, rather than the quality of the work, and if we are not careful, poorly constructed algorithms could remain in use in areas where members of the public do not think there will be knock-on impacts or future problems – C5.

B.3.4 .

whether it's automated or completely automated, or it's assistive or augmenting human decisions, I feel that being completely open and transparent when you're in the very early stages of development tend to be counterproductive, both from a potential IP perspective and also from a public attitude – C3.

B.3.5 .

'if we were to disclose something, the issue is not that people would be worried, but more that there would be an impact on an individual's safety or an impact on the ability of an organisation to target individuals effectively…I would be more than happy to give a strong voice as to why that shouldn't be disclosed, but only if it was from the perspective of significantly impeding the ability of an organisation to safeguard and prosecute – L10.

B.3.6 .

it feels like this would be one of those areas where you could have a effectively a covert version of the standard so that we've gone through the necessary steps and with exactly the same rigour, but it just simply isn't communicated – L7.

B.3.7 .

maybe different levels of information being shared from the point of view of transparency for those different audiences because it wouldn't necessarily be appropriate to share certain police methods with the general public because this could undermine their efficacy and things like that. So I see two different levels and types of transparency – C2.

B.4 Innovation and Commercial Sensitivities

B.4.1 .

The barrier might be that people stop designing tools because they can't be bothered to go through it, which would probably be a bad thing given what we're coming across in [ ] where the new tool will be an awful lot better than what currently exists – C5.

B.4.2 .

I suppose very tight regulation and very high levels of regulation could potentially stifle some of the innovative work that could happen and also potentially be a bar to entry for some small startup companies because it could actually become very expensive to be able to comply – C4.

B.4.3 .

it's a joint responsibility to complete the standard between the customer [and the supplier]" with "slightly more on the customer than the supplier." The supplier can provide a "technical description how it works but I think so much of the use case and the data and the application and how decisions are taken and supported" are on the customer – C3.

B.4.4 .

the technology companies can create some really complex, insightful products that go beyond the understanding of the customers and the forces and I think in doing that, there's a responsibility almost to help the force understand what they're getting into and to be able to on-board them into that process in as transparent a way as possible – L7.

B.4.5 .

we also want to make sure … [there is clear definition] around actually whose responsibility is this, how does the model, how does an algorithm perform over time for example – C4.

B.4.6 .

Police Digital Service is the group helping police forces as a whole to buy the right stuff so they would be a good body to manage [transparency allocation with the private suppliers]. They will probably have far more information generally available to them and more 'in's' with the private companies.' – L2.

B.5 Explainability, Ethical Scrutiny and Evaluation

B.5.1 .

So if you're somebody who doesn't really know what master data management [an example of technical terminology describing a technology enabled discipline] is, then it's a quite tricky starting point to try and describe what you're doing with it without telling a big, big story… where do we draw the line? – L5.

B.5.2 .

We wanted from the start to bring in approach where we know what it's doing. We know why it's making the decisions it's making, and the cops that are using it will know why it's making the decisions it's making – C5.

B.5.3 .

we don't have a single view of our or people's data which most forces don't, actually. Most medium, large organisations don't. However, the public will expect us to have that and that's what we're trying to do – L5.

B.5.4 .

'A really quick tangible example is that we know that there's cell site data that exists and we wanted to consider that information to look at where mobile phones were pinging on masts and to be able to work out county lines and the movement of drugs and the movement of victims connected to that. We didn't go down that route because we knew that ethically, you know, adhering to all of the submissions and the papers and the associated discussions, it was probably going to be vetoed and put to one side so that that stifled our innovation – C6.

B.5.5 .

I would be a little bit wary about disclosing the development of something before it's actually been decided that it will be operationally used, before it's gone through the internal ethical considerations about whether it's appropriate – C2.

B.5.6 .

Our interest is ensuring that we are in the conversation about how to make sure [algorithmic systems are] understood in terms of any biases or anything that might be unacceptable in terms of how we process data. This is for us part of engaging in that conversation and with something like the algorithmics standard – C1.

B.5.7 .

precision and recall rates are defined by various bits and pieces, but ultimately it's a percentage, and when you see that percentage in isolation, you can immediately perceive it to mean something that perhaps it doesn't. So if the precision or recall rate is 60%, some people might think '60% doesn't sound very good at all. That means that 40% is inaccurate or it's not working', whereas actually, if there's some kind of industry standard or expectation that gives the reader an opportunity to benchmark that [figure] against what is roundly perceived as acceptable, and exceeding expectations. The context in which that is done is also equally important – L10. (...)

B.6 Resourcing and Implementation Concerns

B.6.1 .

Depending on what else is happening, for example if there was a big event, there would be change freezes everywhere and ...people being pulled in all sorts of different directions, so if it coincides with a big event, for example, you might just be saying…'No, we're sorry, we just can't do this – L3.

C SCOPE OF THE STANDARD DRAFT COMPARED TO POLICING USE CASES

At the point of doing the interviews, an algorithmic tool was defined within the Standard as any 'product, application, or device that supports or solves a specific problem, using complex algorithms'. The guidance stated that tools will potentially be in scope if they involve one or more of the following:

- complex statistical analysis, complex data analytics, or machine learning - for example neural nets or deep learning;
- potential public effect;
- replace or assist human decision-making.

The full definitions of the draft standard can be found below in Appendix D.

To **illustrate the challenges of deciding whether a specific tool is currently within the scope of the Standard (or within the priority scope), we have developed the following scoring mechanism based on the existing definitions.** According to the scope, all algorithmic tools are generally in scope, but there is a narrower definition of priority scope. We can score algorithmic tools according to the following:

Tools that:

(1) Engage directly with the public – 3 points.
(2) Involve:
  A Complex statistical analysis – 1 point.
  B Complex data analytics – 1 point.
  C Machine learning – 1 point.
(3) A Has a potential legal, economic, or similar impact on individuals or populations – 1 point.
  B Affects procedural or substantive rights – 1 point.
  C Affects eligibility, receipt or denial of a programme – 1 point.
(4) A Replaces human decision making – 1 point.
  B Assists or adds to human decision making – 1 point.

To be included in the priority scope, a tool needs to score 3 points or above, either all from 1, or at least one point from each 2, 3, and 4.

Additional definitions not included in the Standard:

- The standard does not define 'complex' with respect to complex algorithms. In this analysis, we define 'complex' as an action that cannot be reasonably performed by a person in less than a minute.
- We consider any application where the rules of the algorithm are derived from the data as machine learning.

General observations:

- The vast majority of operational tools will receive a point from 4.a or 4.b, while general use tools, e.g., internet browsers, mail clients, etc., will not. It is unclear whether tools used in supporting divisions such as HR, maintenance, etc, should receive a point from 4.
- All tools that may inadvertently contain biases should receive a point from 3.a. However, it is not clear if this will be interpreted in this way by all who read the Standard. This means that the majority of operational tools will receive at least one point from 3 as well.
- Following the above, the decision whether a tool falls within the priority scope will be decided by 2, i.e., will depend on the complexity of the analysis, which is not currently well defined within the scope.

To highlight what is caught by the current scope, we give illustrative examples of algorithmic tools and analyse whether or not they fall within the scope of the Standard. According to the scope, all algorithmic tools are generally in scope, but there is a narrower definition of 'priority scope'. We will score the algorithmic tools according to rules presented above. To be included in the priority scope, a tool needs to score 3 points or above, either all from 1, or at least one point from each 2, 3, and 4.

C.1 Use case examples

- **Crime mapping software:**
  - Visualization of past data only – if the tool only shows points on the map where past crimes (of a certain type) have occurred, it will likely fall within the broader definition of the scope but not within the priority scope.
  - 'Hot spot' prediction – where data on past crime is used to predict where and when future crime will occur will likely be within the priority scope:
    * 1 point from 2.c as these tools are likely to utilise machine learning.
    * 1 point from 3.a because the prediction of hotspots in certain neighbourhoods may impact the level of policing in those areas, impacting the local population.
    * 1 point from 4.a as these tools are likely to assist in resource allocation and patrol planning.
  - Mapping software that includes data processing but does not make predictions may be included in the priority scope if 2.a, 2.b, or 2.c applies.
- **Data infrastructure software:**
  - May fall inside the broad category but are unlikely to fall within the priority scope due to 2, unless:

    * The tool contains a 'link matching' component, i.e., data from two or more sources are combined and records are matched in an automated way.
    * Any type of automated score calculation occurs within the tool, e.g., risk of violence associated with an individual.
- **Data analytics suites:**
  - These types of tools often contain a range of functions within one tool, ranging from data-filtering to mapping, and sometimes assigning risk scores to individuals or places. If one of the functionalities is within scope, it is not clear whether the whole tool suite is then considered within the scope or just the specific functionality.
- **Individualised risk scores:**
  - Algorithmic tools used to produce risk scores will often use machine learning to generate the predications, ensuring they fall within the priority scope. However, this may not always be the case. An individual may be assigned a risk score based on simple rules, in which case these tools will fall outside the scope. As 'complex' is not well defined in the scope, this may be down to the police force's discretion.
- **Solvability tools:**
  - As above, based on the current definition it will depend on whether or not the analysis is considered 'complex'.
- **Facial recognition tools:**
  - Falls within the priority scope as they use machine learning (2.c), may be biased (3.a), and assist human decision making (4.c).
- **Chatbots:**
  - Chatbots will be included in the priority scope as they interact directly with the public (3 points from 1).

D THE DEFINITION OF SCOPE OF THE DRAFT STANDARD

As the definition of scope has changed, in part due to the results of this study, we provide a print of the webpage that contained the definition at the time of the pilot and subsequent interviews. Please see the following 2 pages below.

GOV.UK

Guidance

Provide information on how you use algorithmic tools to support decisions (pilot version)

Follow this guidance if you're a government or public sector employee using algorithmic tools to support decisions in your organisation.

From:
Central Digital and Data Office (/web/20220706195947/https://www.gov.uk/government/organisations/central-digital-and-data-office)
Published
29 November 2021
Last updated
6 December 2021 —

Contents

Use this CDDO template to provide information about the algorithmic tools you use, and why you're using them.

After you provide your information, we'll:

- publish it in the Algorithmic Transparency Standard collection (https://web.archive.org/web/20220706195947/https://www.gov.uk/government/collections/algorithmic-transparency-standard), so people can see the algorithmic tools you use and why you're using them
- reformat your information into the algorithmic transparency data standard (https://web.archive.org/web/20220706195947/https://www.gov.uk/government/publications/algorithmic-transparency-data-standard)

This will help people who use, regulate, or are affected by the results of your algorithmic-assisted decisions.

The Algorithmic Transparency Standard is part of the government's National Data Strategy (https://web.archive.org/web/20220706195947/https://www.gov.uk/government/publications/uk-national-data-strategy). Find out more about the Algorithmic Transparency Standard (https://web.archive.org/web/20220706195947/https://www.gov.uk/government/collections/algorithmic-transparency-standard).

This is the first version of the template and we'll be iterating it based on your feedback. You can submit your suggestions to help develop the template by emailing data-ethics@digital.cabinet-office.gov.uk (https://web.archive.org/web/20220706195947/mailto:data-ethics@digital.cabinet-office.gov.uk) before 28 February 2022.

1. Check if your tool is in scope

We encourage you to provide information about all the algorithmic tools you're using. However, in the initial phase of the Algorithmic Transparency Standard, we'll prioritise publishing information about tools that either:

- engage directly with the public - for example a chatbot
- meet at least one criteria in each of the three areas below

Related content

- Algorithmic transparency data standard (/web/20220706195947/https://www.gov.uk/government/publications/algorithmic-transparency-data-standard)
- Algorithmic transparency template (/web/20220706195947/https://www.gov.uk/government/publications/algorithmic-transparency-template)
- Algorithmic Transparency Standard (/web/20220706195947/https://www.gov.uk/government/collections/algorithmic-transparency-standard)

Collection

- Algorithmic Transparency Standard (/web/20220706195947/https://www.gov.uk/government/collections/algorithmic-transparency-standard)

You should prioritise providing information about tools that have a legal, economic or similar impact on individuals, and replace human decision making.

You can read definitions of the terms we use (https://web.archive.org/web/20220706195947/https://www.gov.uk/government/collections/algorithmic-transparency-standard#definitions).

Technical specifications

Your tool will be in scope if involves one of the following:

- complex statistical analysis
- complex data analytics
- machine learning - for example neural nets or deep learning

Potential public effect

Your tool will be in scope if it does one of the following:

- has a potential legal, economic, or similar impact on individuals or populations
- affects procedural or substantive rights
- affects eligibility, receipt or denial of a programme - for example receiving benefits

Impact on your decisions

Your tool will be in scope if it does one of the following:

- replaces human decision making
- assists or adds to human decision making - for example it provides evidence for decisions

2. Fill in the template

Fill in the algorithmic transparency template (https://web.archive.org/web/20220706195947/https://www.gov.uk/government/publications/algorithmic-transparency-template) using the guidance on the template.

3. Send your completed template to us

Send your completed template to data-ethics@digital.cabinet-office.gov.uk (https://web.archive.org/web/20220706195947/mailto:data-ethics@digital.cabinet-office.gov.uk).

We'll then help you with next steps, and publish it in the Algorithmic Transparency Standard (https://web.archive.org/web/20220706195947/https://www.gov.uk/government/collections/algorithmic-transparency-standard) collection.

Published 29 November 2021
Last updated 6 December 2021 + show all updates

1. 6 December 2021
   Added deadline for suggestions to help develop the template.
2. 29 November 2021
   First published.