

A Sociotechnical Audit: Assessing Police Use of Facial Recognition

Evani Radiya-Dixit
er595@cam.ac.uk
Minderoo Centre for Technology and Democracy,
University of Cambridge
Cambridge, UK

Gina Neff
gsn23@cam.ac.uk
Minderoo Centre for Technology and Democracy,
University of Cambridge
Cambridge, UK

ABSTRACT

Algorithmic audits are increasingly used to hold people accountable for the algorithms they implement. However, much work remains to integrate ethical and legal evaluations of how algorithms are used into audits. In this paper, we present a sociotechnical audit to help external stakeholders evaluate the ethics and legality of police use of facial recognition technology. We developed this audit for the specific legal context of England and Wales, and to bring attention to broader concerns such as whether police consult affected communities and comply with human rights law. To design this audit, we compiled ethical and legal standards for governing facial recognition, based on existing literature and feedback from academia, government, civil society, and police organizations. We then applied the resulting audit tool to three facial recognition deployments by police forces in the UK and found that all three failed to meet these standards. Developing this audit helps us provide insights to researchers in designing their own sociotechnical audits, specifically how audits shift power, how to make audits context-specific, how audits reveal what is not transparent, and how audits lead to accountability.

CCS CONCEPTS

• **Social and professional topics** → **Surveillance**; **Technology audits**; • **Security and privacy** → **Human and societal aspects of security and privacy**.

KEYWORDS

algorithmic audits, accountability, ethical and legal considerations, facial recognition technology

ACM Reference Format:
Evani Radiya-Dixit and Gina Neff. 2023. A Sociotechnical Audit: Assessing Police Use of Facial Recognition. In *2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23), June 12–15, 2023, Chicago, IL, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3593013.3594084

1 INTRODUCTION

With the growing adoption of technology, algorithmic audits are increasingly used to hold people accountable for the algorithms they implement. Journalists, regulators, academics, and others have used audits to analyze the biased outcomes of algorithmic systems. Algorithms can reflect and perpetuate existing power dynamics in society, and audits help interrogate and understand these dynamics [65]. For example, the Gender Shades audit revealed how facial analysis algorithms fail to recognize dark-skinned individuals [21], which points to structures that render Black people invisible in society [13].

Algorithmic audits have exposed biased outcomes in a variety of domains including criminal justice [6, 40, 89], tenant screening [56, 83], healthcare [31, 72], and online advertising [57, 92]. Many audits have helped hold decision-makers responsible for the impacts of algorithms, motivating moratoriums [50], lawsuits [91], and regulation on the use of technology [87].

Although audit studies have had significant impact, there remains much work to integrate ethical and legal evaluation into audits. As the FAccT community has shown, understanding algorithmic power requires understanding the power structures in which algorithmic systems operate [12, 18, 55, 58, 76, 81, 85, 98]. Broadening the scope and impact of algorithmic audits by considering ethical and legal issues is challenging. Consider, for example, whether the use of an algorithm complies with data privacy law or human rights law. Such work has to be anchored in different legal jurisdictions and traditions. The same can hold for important ethical principles of fairness, transparency, and accountability. For example, do the developers of an algorithm engage with affected communities for feedback, or are they held accountable if people are harmed? Such questions are often highly dependent on the context or the type of system being audited.

In this paper, we present an external audit that evaluates the ethics and legality of police use of facial recognition technology (FRT) in England and Wales. In the debate on facial recognition, there has been much attention on algorithmic bias. Part of the challenge motivating our work was that even so-called bias-free facial recognition tools can still be used in ways that harm and discriminate against marginalized communities [4, 17, 36]. As a tool of surveillance, FRT can pose threats to privacy, the right to protest, and other civil liberties, especially in the U.S. and UK for Black bodies [19]. Thus, how FRT is used has broad implications for equality and accountability that go far beyond notions of algorithmic bias.

Our audit scorecard (Section 5) establishes a set of ethical and legal standards for governing FRT. We developed this audit using existing research literature and soliciting input from academia, government, civil society, and police organizations. The audit evaluates (1) how police show compliance with the law, (2) how reliably FRT performs in practice, (3) how the use of FRT shifts police decisions, and (4) how much expertise and oversight exists over police use of the technology. By addressing such questions, our audit helps anchor police use of FRT in a much larger sociotechnical context and helps people examine the politics, ethics, and legality of FRT.



This work is licensed under a Creative Commons Attribution International 4.0 License.

FAccT '23, June 12–15, 2023, Chicago, IL, USA
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0192-4/23/06.
https://doi.org/10.1145/3593013.3594084

We applied our audit to three cases of FRT deployments in the real world (Section 6). The first was the *Bridges* court case. From 2017 to 2019, South Wales Police conducted operational trial deployments of live FRT, which were later ruled unlawful in *R (Bridges) v. Chief Constable of South Wales Police*. The second was the London Metropolitan Police Service's operational trial deployments of live FRT from 2016 to 2019. The third was South Wales Police's three-month operational trial of FRT using a mobile phone application from 2021 to 2022. We found that these three deployments failed to meet the ethical and legal standards that we compiled in the audit.

While we aimed to evaluate how FRT was being used with this sociotechnical audit, performing well on the audit does not green-light the use of FRT. Rather, the audit helps reveal the risks of FRT, evaluate legal compliance, and advance policy and oversight options that help redress real harms that people experience.

Our contribution with this paper is two-fold: we demonstrate our specific audit tool, and also distill key insights for how to design sociotechnical audits more generally (Section 7). In particular, we discuss how audits shift power, how to make audits context-specific, how audits reveal what is not transparent, and how audits lead to accountability. Our work can help expose societal harms and improve accountability in how algorithmic systems are used.

2 BACKGROUND

2.1 Key Definitions

An audit is a tool for analysis or inspection, often to evaluate the compliance of a system with respect to predefined standards [1, 65, 79]. The audit that we present here evaluates the sociotechnical system of police use of facial recognition technology in England and Wales. The term sociotechnical refers to the interactions between people and technology [27].

Facial recognition technology refers to a digital tool used to perform tasks on images or videos of human faces [22]. This audit extends to all types of face identification or one-to-many facial recognition tools. Here, a facial image or probe image is first captured and then compared with a database or watchlist of known facial images in order to determine if there is a match. One type of face identification technology is live FRT, where images, such as from a live camera feed, are compared to the watchlist in real time. In contrast, retrospective FRT involves images, such as from surveillance camera footage, being compared to the watchlist at a later point in time. Finally, mobile phone or operator initiated FRT refers to when images, captured using a mobile phone, are compared to the watchlist in near real time. Details of FRT performance metrics in the context of policing are available [75].

Our audit evaluates police use of FRT with respect to legal and ethical standards. Legality is defined as compliance with the law in England and Wales. Currently, there is no explicit legal basis for police use of FRT in the UK. Thus, legal standards in the audit are primarily informed by the Human Rights Act 1998, the Equality Act 2010, and the Data Protection Act 2018, which are relevant as FRT interferes with rights protected by these acts.

We use the term ethics to encapsulate the principles of fairness, transparency, and accountability. We refer to fairness as the elimination of the discriminatory effects of police use of FRT on people. Transparency is the quality of police being open about their use of FRT in a complete, understandable, and accessible manner. Accountability refers to the state of the police being responsible or answerable to the public for the societal impacts of their use of FRT.

2.2 Harms of Police Use of Facial Recognition

Police often advocate for the adoption of FRT to help address crime and identify vulnerable, missing, or wanted individuals. However, the use of FRT can pose serious threats to fundamental rights and disproportionately impact marginalized groups.

FRT interferes with the rights to privacy and data protection through the use of personal data, often without people's knowledge or consent. Police use of this technology also impacts the rights to free expression and assembly. FRT surveillance can inhibit our ability to express ideas and generate a "chilling effect" where people withhold from exercising their rights, such as the right to protest, out of fear of the consequences [45].

FRT adoption also has serious implications for equality. Historically, surveillance systems have been used to monitor marginalized groups [13, 19]. Police use of these systems can perpetuate disproportionate policing practices that often target Black and low-income communities in the UK [11, 39, 64, 101]. Additionally, studies have shown that FRT disproportionately misidentifies women, people of color, and people with disabilities [21, 23]. An incorrect identification can lead to unwarranted police intrusions (e.g., fingerprinting, stop and search, or wrongful arrest). While reducing the bias in FRT may mitigate some harms, it does not eliminate the harms that come from its discriminatory use and impact on human rights.

Moreover, police use of FRT can perpetuate existing harms within the criminal justice system. Historically, there have been historical issues of racism, misogyny, and over-policing of marginalized communities in the UK that continue today [3, 53, 61, 62, 84]. While surveillance systems such as FRT are often justified as tools for security, they can threaten the safety of vulnerable communities [63]. Instead of using FRT surveillance to address crime, many people advocate for addressing underlying inequities and investing in education, housing, and community welfare [13, 33, 34, 97].

2.3 Related Work

Below we discuss related frameworks for assessing algorithmic systems, highlighting examples of research on FRT.

Algorithmic systems can perpetuate existing systems of oppression that cause real-world harm in society. Audits are powerful tools to help understand these harms and hold entities accountable [20, 96]. A **technical audit** evaluates the technical elements (e.g., the outputs) of an algorithmic system to reveal harmful behavior [9]. Technical audits have exposed bias and discrimination in search [71], recommendation [43], and language processing [86]. In computer vision, several audits show that FRT discriminates on the basis of race and gender [10, 21, 59].

Taking a sociotechnical view, an audit can also examine the interplay between people and the algorithm [60]. For example, the Gender Shades technical audit was followed by a sociotechnical audit of how public pressure influenced companies to address bias in their FRT systems [77].

A **sociotechnical audit** evaluates the human and technical elements of an algorithmic system to uncover harms [60]. Some

sociotechnical audits ask important questions about how people are impacted by algorithmic systems [52, 58]. Others evaluate users' experiences of algorithms in particular domains like online services and advertising [60, 68]. Sociotechnical audits can also integrate ethical and legal standards [1, 35]. In the U.S. context, researchers developed a scorecard with criteria on the ethics and legality of law enforcement use of FRT [46]. Such criteria are highly dependent on the context and jurisdiction in which FRT is used. To the best of our knowledge, there is no set of ethical and legal criteria developed to externally evaluate police FRT deployments in the UK, and our audit helps fill this gap.

Typically, audits are conducted after the implementation of an algorithmic system. Audits complement **impact assessments** that analyze the possible consequences of an algorithmic system, usually *before* implementation [1, 66, 95]. While impact assessments are often executed internally, audits like ours can help provide outsider oversight and actual accountability [80].

3 MOTIVATION

Researchers in algorithmic auditing suggest that the purpose of an audit is to reveal blind spots rather than to authorize the use of a technology [78]. Thus, performing well on this audit does not green-light FRT adoption nor carry enough weight to overturn an existing moratorium. Rather, the audit can help assess whether ethical and legal standards to mitigate harm are met.

Our audit can reveal deficiencies such as an inadequate legal basis, lack of community oversight, or discriminatory use of FRT. The audit could also help evaluate legal compliance. The audit is contextualized for the England and Wales jurisdiction and is informed by the Human Rights Act 1998, the Equality Act 2010, the Data Protection Act 2018, and national guidance from the UK government. Satisfying the audit does not mean that the police comply with the law. Rather, the audit points to legal risks that need to be considered. Finally, the audit can inform policy, advocacy, and oversight on the use of FRT and lead to greater accountability.

For the audit to be meaningful, it should be administered by an entity independent of the police. Auditor independence is crucial to mitigate conflicts of interest that could yield biased audit results [67, 80]. Key external stakeholders who might administer this audit or use the findings include:

- **Regulatory bodies** can use the audit to monitor and enforce the law for police use of FRT, administer inspections into how police are using FRT, and provide national guidance.
- **Oversight bodies** can use the audit to administer inspections into how police are using FRT, provide ethics scrutiny on the use of FRT, and improve public understanding of FRT.
- **Policymakers** can use the audit to inform debates, inquiries, and legislation on police use of FRT.
- **Civil society** can use the audit to campaign for policies, pursue strategic litigation that challenges police use of FRT, and provide expert evidence on FRT to government bodies.
- **The public**, especially impacted individuals or parties acting on their behalf, can use the audit to understand how police are using FRT and seek remedy for any resulting harm.

The audit can be conducted after a police force's FRT deployment(s). Any evaluation using this audit should be based on information that is known and accessible to the public. This helps to assess how transparent police forces are with the public. Additionally, publishing key audit results can mobilize change in whether and how police implement FRT [67]. We discuss the limitations of disclosing audit results in Section 8.

4 METHODOLOGY

To construct this audit, we translated high-level principles of ethics and legality into what they mean for practice within the context of police use of FRT in England and Wales. As described in Section 2.1, we define ethics by the principles of fairness, transparency, and accountability and legality as compliance with the law in England and Wales. When designing audit questions, we followed standards in social science survey research [8, 24, 70]. Here, we detail how we used various sources to move from principles to practice and trace an example audit question to illustrate this process.

We began with what ethics and legality mean in the general context of public sector use of data and artificial intelligence (AI) systems. Here, we used frameworks developed by the UK government that reflect the types of questions the government expects public agencies including police forces to answer. Specifically, we used the Data Ethics Framework [25], the Guide to Using AI in the Public Sector [26], and the Algorithmic Transparency Standard [38]. By consolidating and grouping questions from these sources, we arrived at an initial draft of the audit. For example, based on the Data Ethics Framework, one audit question was: "What are the governance mechanisms that enable domain experts to challenge the FRT project?"

We then adapted this initial draft to the specific context of police use of FRT in England and Wales using existing research literature. We drew on documents focused on FRT, surveillance, personal data, and policing technologies. We revised the general audit questions based on documents from various perspectives:

- **Users:** We examined documents on FRT developed by police forces in England and Wales to understand the current landscape and gaps in how police are using the technology.
- **Courts:** We drew on legal challenges to police use of FRT and related court cases to gather perspectives from courts that interpret the laws.
- **Legislators:** We used reports developed by UK legislative committees on policing technologies.
- **Regulators:** We incorporated guidance developed by UK regulatory bodies on FRT and data protection.
- **Academia:** We drew on academic evaluations of police use of FRT in England and Wales to understand known ethical and legal issues that have arisen in past FRT deployments.
- **Advisors:** We leveraged resources on data and AI usage developed by oversight and advisory bodies such as local ethics committees, professional bodies, and government entities.
- **Auditors:** We examined evaluations conducted by algorithmic auditors to test the performance of FRT systems.
- **Civil society:** We used resources on FRT and AI governance developed by civil society groups focused on protecting privacy, human rights, and civil liberties in the digital age.

Details of the specific sources used to generate each audit question are available [75]. For example, reports from advisory and legislative bodies discuss how local ethics committees are an existing governance mechanism to oversee policing technologies in England and Wales [16, 54]. Thus, we revised the aforementioned audit question to: "Are there clear processes for the *ethics committee* to influence if and how FRT is implemented?"

We then revised the audit based on informal feedback from 35 stakeholders from police bodies, government, academia, and civil society. Engaging with civil society helped us make our audit questions on human rights, accountability, and oversight more critical. For example, we strengthened the previously mentioned audit question to: "Are there clear processes for the ethics committee to influence if and how FRT is implemented, including the *power of veto* for the FRT project?" We also spoke with police organizations, government bodies, and ethics committees, which improved the practicality of the audit and grounded it in the current landscape of police use of FRT. Engaging with legal scholars helped us add detail to legal questions in the audit based on human rights law, data protection law, and case law on surveillance systems. Finally, many stakeholders pointed us to research and advocacy efforts to examine policing technologies, and we adapted our audit to build on these ongoing efforts. Stakeholders who agreed to be acknowledged are included in the Acknowledgements.

This audit is composed of yes/no questions that are scored with either 1 (yes) or 0 (no), alongside an explanation.[1] While the qualitative explanation captures nuances, the quantitative scoring system helps simplify the audit, compare results across police forces, and replicate results with different auditors. Additionally, questions are grouped by different aspects of ethics and legality, such as data protection and non-discrimination. The scores in each group are added to produce a composite measure or index for the given aspect. Certain concepts (e.g., data protection) are multi-dimensional, and each sub-component of the index reflects a distinct dimension [8, 24, 70]. Using indexes helps summarize the results and identify where there are deficiencies.

We applied this audit to three case studies: (1) *Bridges* case on South Wales Police's use of live FRT, (2) Metropolitan Police Service's use of live FRT, and (3) South Wales Police's use of mobile phone FRT. We choose these cases based on: (a) notability of the cases, (b) a sample of different police forces, and (c) different types of FRT being used. When we applied the audit, we identified gaps and ambiguities in some questions and subsequently refined them.

5 SOCIOTECHNICAL AUDIT SCORECARD

We present this sociotechnical audit as a tool to assess the ethics and legality of police use of FRT. Contextualized for England and Wales, the audit comprises four sections. The Legal Standards section evaluates how police show legal compliance for the use of FRT. The Technical Reliability section evaluates how reliably FRT performs in practice. The Human Decision-Making section evaluates how the use of FRT shifts police decisions. Finally, the Expertise and Oversight section evaluates how much expertise and oversight exists over police use of FRT. Detailed explanations of the audit sections are available [75].

This audit should be used to reveal the harms of FRT. The questions are not exhaustive and not to be treated as a checklist. Each yes/no question is scored with either 1 (yes) or 0 (no), accompanied by an explanation. If the answer is unknown or inaccessible to the public, the question is scored with 0 (no). Each subsection is then scored by the number of questions within it that scored 1 (yes).

5.1 Legal Standards

In Accordance with the Law (Human Rights Act 1998)

a. Are there clear, objective, and limited criteria for who can be included in the watchlist, including with regard to the image source and the seriousness of offense or risk?
b. Are there clear, objective, and limited criteria for where and when FRT can be used, including mandating reasonable suspicion that persons on the watchlist will be at the location and requiring a high grade of intelligence for the police intelligence case that supports FRT use?
c. Are there clear, objective, and limited criteria concerning third-party access to the data collected or retained, including with regard to what data can be shared, with whom it can be shared, and for what specific purpose it can be shared?

Score: / 3

Necessary in a Democratic Society (Human Rights Act 1998)

d. Have police identified less intrusive alternative measures and proven that FRT is strictly necessary compared to these measures using scientifically verifiable evidence?
e. Have police conducted distinct necessity tests with an evidence-based justification for each category of individuals on the watchlist?
f. Have police shown that FRT does not disproportionately limit the human rights of affected persons, including those who are misidentified, not on the watchlist, or impacted by unwarranted intrusions?

Score: / 3

Data Protection (Data Protection Act 2018)

g. Before using FRT, have police carried out and published a data protection impact assessment and appropriate policy document for sensitive data processing?
h. Beyond social media or website publishing, have police used other means to inform potential data subjects or most people in their jurisdiction in advance about when, where, why, and how FRT is being used and how they can exercise their individual rights?
i. Are there clear measures to ensure data subjects can exercise their individual rights including the rights to rectification, erasure, and object with clear justifications if exemptions apply?
j. Do police check the watchlist against the data source close to the time of deployment to ensure the watchlist is accurate and up to date?
k. Are there clear measures to ensure that watchlist images are lawfully held, have a known provenance, and exclude unconvicted custody images?
l. Via direct consultation, have police proactively considered views of the public, especially marginalized communities, on the particular type of FRT and justified a disregard of the views if relevant?
m. Have police published their procurement contracts and data-sharing agreements with other parties?

Score: / 7

[1] This scoring mechanism was inspired by the non-profit White Coats for Black Lives' Racial Justice Report Card [100]. We use a similar design for our sociotechnical audit scorecard.

Non-Discrimination (Human Rights Act 1998 and Equality Act 2010)

n. Before using FRT, have police carried out and published an equality impact assessment?
o. For each deployment, have police published the demographic makeup of the watchlist?
p. For each deployment, have police published the demographic makeup of the population where FRT is used?
q. For each deployment, have police published the demographic data for arrests, stop and searches, and other outcomes resulting from the use of FRT?

Score: / 4

Free Expression and Assembly (Human Rights Act 1998)

r. Have police assessed FRT's potential "chilling effect" on the rights to freedom of expression and assembly to inform the legal test of "necessary in a democratic society"?
s. Do police preclude using FRT to identify those peacefully participating in an assembly?

Score: / 2

5.2 Technical Reliability

Algorithmic Fairness (Equality Act 2010)

a. Before using FRT, have police evaluated and published the demographic makeup of the training dataset to ensure the dataset is representative of the population where it is to be used?
b. Before using FRT, have police evaluated and published FRT's performance across demographic groups, in different conditions that match FRT's operational use, to ensure FRT performs well and similarly across the population?

Score: / 2

Robust Practice (Data Protection Act 2018)

c. Are there safeguards precluding the use of FRT with an unsuitable low-quality probe or watchlist image?
d. Have police pre-established and met thresholds for the FRT system's accuracy (precision, false positive rate, true positive rate) to inform the legal test of strict necessity for personal data processing?

Score: / 2

Deployment Performance (Equality Act 2010)

e. Does FRT perform well (precision, false positive rate, true positive rate) and similarly across demographic groups?

Score: / 1

5.3 Human Decision-Making

Human Review

a. Is there a transparent evaluation that shows human review of the FRT matches is reliable, given the accuracy of officer-verified matches and the amount of time an officer has to review an FRT match?

Score: / 1

Preparation

b. Is training for the particular type of FRT mandated for police officers using the technology?
c. Are there clear standards for technical training on using FRT, data protection training, and training on risks including differential treatment, function creep, and unwarranted intrusions?
d. Has there been a documented non-operational research trial of FRT with informed consent from participants before the operational use of FRT for policing?

Score: / 3

Accountability

e. Are there clear measures for police to document cases of harm resulting from the use of FRT such as differential treatment, function creep, or unwarranted intrusions?
f. Do police have a whistleblower protection policy to protect persons who reveal FRT misuse?
g. Is there a clear redress mechanism (beyond judicial review and usual complaint procedures) for harmed individuals and groups to participate in an investigation into police use of FRT?
h. Are there clear measures to ensure that the redress mechanism is procedurally fair?

Score: / 4

5.4 Expertise and Oversight

Ethics Committee

a. Is regular oversight from an ethics committee mandated throughout the life of the FRT project?
b. Are there clear processes for the committee to influence if and how FRT is implemented, including the power of veto for the FRT project?
c. Is the committee an independent body from police organizations with members having non-policing backgrounds and with safeguards to ensure the committee's sustainability even without political support?
d. Is the committee diverse in terms of demographic makeup and independent expertise in human rights, equality, and data protection?
e. Are detailed meeting minutes published, including briefing papers, discussions, and conclusions?

Score: / 5

Civil Society and Experts

f. Are there transparent, proactive consultations with civil society and independent experts on the particular type of FRT?
g. Are police required to consider the advice from consultations and transparently explain the outcomes, including providing a justification if the advice is not followed?

Score: / 2

Community Engagement

h. Are there clear, proactive processes for the public, especially marginalized communities, to influence if and how FRT is implemented?
i. Are all FRT materials accessible to people with disabilities and provided in immigrant languages?

Score: / 2

Table 1: Summary of the audit scorecard for the *Bridges* case on South Wales Police's trial of live facial recognition (LFR).

Metric	Score	Notes
1. Legal Standards		
In accordance with the law	0 / 3	Lack of clear limits for watchlist, usage, and data access
Necessary in a democratic society	0 / 3	Inadequate necessity and proportionality assessments
Data protection	2 / 7	Up-to-date watchlist, but inadequate measures to ensure rights
Non-discrimination	1 / 4	No published demographic data for watchlist, usage, and arrests
Free expression and assembly	0 / 2	No assessment of chilling effect; no limit on LFR at protests
2. Technical Reliability		
Algorithmic fairness	0 / 2	No evaluation of LFR's data bias or algorithmic bias
Robust practice	0 / 2	Low-quality images could be used; no pre-established thresholds
Deployment performance	0 / 1	Poor LFR precision of 24%; different accuracy across gender
3. Human Decision-Making		
Human review	0 / 1	Human review of LFR-generated matches had 69% precision
Preparation	0 / 3	Only technical training; lack of training for initial deployments
Accountability	0 / 4	Whistleblower policy only created in 2019; lack of redress for harms
4. Expertise and Oversight		
Ethics committee	0 / 5	Lack of regular oversight; lack of diversity and independence
Civil society and experts	0 / 2	Lack of proactive and effective consultations on LFR use
Community engagement	0 / 2	Lack of community oversight; lack of accessible documents

6 SOCIOTECHNICAL AUDIT CASE STUDIES

In July 2022, we applied this audit to three facial recognition deployments by police in England and Wales. We found that these deployments lacked (a) evidence of a lawful interference with privacy rights, (b) transparent evaluations of discrimination, (c) measures for remedy for harmed persons, and (d) regular oversight from an independent ethics body and affected communities.

In this section, we summarize each case study. Full case studies with explanations of how each audit question was scored are available [75].

6.1 *Bridges* Case on South Wales Police's Trial of Live Facial Recognition

Our first case is the operational trial deployments of live facial recognition (LFR) conducted by South Wales Police (SWP) from May 2017 to April 2019. In *R (Bridges) v. Chief Constable of South Wales Police*, the Court of Appeal ruled that these deployments were unlawful as "there was no clear guidance on where [LFR] could be used and who could be put on a watchlist, a data protection impact assessment was deficient and the force did not take reasonable steps to find out if the software had a racial or gender bias" [41, 82]. As shown by the scorecard summary in Table 1, our sociotechnical audit revealed additional legal and ethical concerns beyond the scope of the court case.

First, SWP did not establish limits on the use of LFR at assemblies. In fact, the technology was used at a peaceful anti-arms protest [7, 14], interfering with the human rights to freedom of expression and assembly, without evidence that the legal requirement of "necessary in a democratic society" was met. SWP's data protection impact assessment and policy documents did not acknowledge nor address LFR's impact on the rights to freedom of expression and assembly.

Second, LFR does not perform well or similarly across demographic groups. Out of the matches that LFR generated, only 24% were verifiably correct. There was also a higher false positive rate for women (82%) compared to men (66%). This raises serious concerns that people faced unwarranted police interventions due to misidentifications.

Additionally, there was a lack of effective oversight over the use of LFR. While SWP had early engagements with the SWP Joint Independent Ethics Committee, regular and transparent oversight was not provided throughout the lifecycle of the LFR project. During committee meetings, there were no independent experts in human rights, equality, or data protection in attendance, even though such expertise has been documented as crucial for the oversight of technologies such as LFR [54, 74, 93, 99].

Moreover, there remained concerns about the committee's independence. Although there were some independent members, the committee also included police officers and is a body situated within the police force. In fact, during meetings, 63% of attendees were members of SWP and 71% were members of either SWP or the South Wales Police and Crime Commissioner. Finally, there were no consultations with the public, especially marginalized communities, on how and whether LFR was implemented.

Table 2: Summary of the audit scorecard for the Metropolitan Police Service's trial of live facial recognition (LFR).

Metric	Score	Notes
1. Legal Standards		
In accordance with the law	1 / 3	Limits for data access, but lack of limits for watchlist and usage
Necessary in a democratic society	0 / 3	Inadequate necessity and proportionality assessments
Data protection	0 / 7	Issues of inaccurate data; inadequate measures to ensure rights
Non-discrimination	0 / 4	Some demographics provided, but not for arrests and outcomes
Free expression and assembly	0 / 2	No assessment of chilling effect; no limit on LFR at protests
2. Technical Reliability		
Algorithmic fairness	0 / 2	No published evaluation of LFR's data bias or algorithmic bias
Robust practice	0 / 2	Low-quality images could be used; no pre-established thresholds
Deployment performance	0 / 1	Poor LFR precision of 19%; different accuracy across gender
3. Human Decision-Making		
Human review	0 / 1	Human review of LFR-generated matches had 36% precision
Preparation	0 / 3	No mandated training for LFR; lack of non-operational trial
Accountability	0 / 4	Lack of whistleblower protection; lack of redress for harms
4. Expertise and Oversight		
Ethics committee	0 / 5	Lack of oversight from the start; lack of diversity and veto power
Civil society and experts	0 / 2	Lack of proactive and effective consultations on LFR use
Community engagement	0 / 2	Lack of community oversight; lack of accessible documents

6.2 Metropolitan Police Service's Trial of Live Facial Recognition

The next case is the operational trial deployments of live facial recognition (LFR) conducted by the Metropolitan Police Service (MPS) from August 2016 to February 2019 [69]. We built upon a study conducted by University of Essex researchers on the human rights compliance of these trials. Their report concludes that the trials would likely "be held unlawful if challenged before the courts" given the absence of clear guidance on who was included in a watchlist and the failure to establish that LFR was "necessary in a democratic society" as required by human rights law [45]. Our sociotechnical audit revealed additional concerns related to discrimination and oversight, as illustrated in Table 2.

While MPS published some demographic data in their results, they did not record the demographic breakdown for engagements, stop and searches, and arrests resulting from the use of LFR. This makes it hard to evaluate whether LFR perpetuates racial profiling. There was also no published evaluation of racial or gender bias in the LFR software. MPS conducted an internal evaluation but did not disclose the results. This lack of transparency makes it hard for outside entities to assess the comprehensiveness of the evaluation. As we discuss in Section 7.3, this obscurity reveals how power is concentrated in the police and where change needs to be made.

Since the LFR trial has ended, MPS has pointed to an evaluation undertaken by the National Institute of Standards & Technology [49]. However, citing this evaluation can be misleading: the evaluation shows high accuracy, but it was conducted with high-quality standardized images rather than wild images on which LFR was used. In fact, for MPS' trial, only 19% of LFR matches were verifiably correct. This performance is especially concerning given that the same technology used by MPS misidentified and led to wrongful arrests of Black men in the U.S. [2, 32, 51].

Regarding oversight, MPS engaged with the London Policing Ethics Panel. However, transparent oversight did not begin until several deployments rather than starting from the concept stage of the trial. Even though MPS responded to the panel's recommendations, the panel was advisory and MPS was not required to act upon the recommendations. There were also no experts in human rights, equality, or data protection on the panel, even though this is crucial for the oversight of technologies such as LFR.

6.3 South Wales Police's Trial of Mobile Phone Facial Recognition

Our final case is the operational trial of mobile phone or operator initiated facial recognition (OIFR) conducted by South Wales Police (SWP) from December 2021 to March 2022 [90]. SWP provided more documentation about their use of OIFR in comparison with their LFR trial. However, as illustrated in Table 3, significant gaps remain with regard to ethical and legal standards.

First, the watchlist included all SWP custody images with no limits on the seriousness of offense.[2] This broad inclusion raises

[2]Custody images are photographs taken by police when an individual is arrested.

Table 3: Summary of the audit scorecard for South Wales Police's trial of operator initiated facial recognition (OIFR).

Metric	Score	Notes
1. Legal Standards		
In accordance with the law	0 / 3	Lack of limits for data access and the offense type for watchlist
Necessary in a democratic society	0 / 3	Inadequate necessity and proportionality assessments
Data protection	1 / 7	Up to date watchlist, but inadequate measures to ensure rights
Non-discrimination	1 / 4	Some demographics provided, but not for watchlist and arrests
Free expression and assembly	0 / 2	No assessment of chilling effect; no limit on OIFR at protests
2. Technical Reliability		
Algorithmic fairness	0 / 2	Unknown demographics of training and evaluation datasets
Robust practice	0 / 2	Low-quality images could be used; no pre-established thresholds
Deployment performance	1 / 1	OIFR match returned as the top result on every occasion of use
3. Human Decision-Making		
Human review	0 / 1	No published evaluation of the human review of OIFR matches
Preparation	2 / 3	Non-operational trial conducted, but unclear training standards
Accountability	1 / 4	Whistleblower protection, but lack of redress for harms
4. Expertise and Oversight		
Ethics committee	1 / 5	Some oversight provided, but lack of diversity and independence
Civil society and experts	0 / 2	Lack of proactive and effective consultations on OIFR use
Community engagement	0 / 2	Lack of community oversight; lack of accessible documents

concerns about the legal requirement of "necessary in a democratic society", especially whether SWP conducted distinct necessity tests for people with minor offenses and those with serious offenses. Moreover, the watchlist included images of innocent persons who were arrested but unconvicted, despite these images being unlawful to retain [42].

Second, while SWP took proactive steps to evaluate bias and discrimination, there was a lack of full transparency for these evaluations. SWP evaluated OIFR's accuracy before its operational use and found no evidence of algorithmic bias. However, SWP did not publish the demographic distribution of the evaluation dataset, which is crucial to assess bias. Additionally, SWP provided the demographic data for the people on which OIFR was used, but the demographic data for the watchlist and those arrested are unknown.

With regard to oversight, SWP engaged with the SWP Joint Independent Ethics Committee before and after the OIFR trial. However, the committee consists of police officers and is a body situated within the police, raising concerns about the independence of the oversight. Based on the most recently published meeting minutes, there were no independent experts in human rights, equality, or data protection on the committee. Moreover, SWP did not consult the public nor civil society to gather feedback for the OIFR trial.

Finally, across all three case studies, there was no clear framework to ensure accountability for the misuse or failure of FRT [54]. There was a lack of robust redress mechanisms for those harmed by FRT deployments. Additionally, police force documents were not fully accessible to people with disabilities or provided in immigrant languages. This lack of accessibility makes it difficult for certain groups to understand how FRT impacts them and to seek remedy in the case of harm.

7 DESIGNING SOCIOTECHNICAL AUDITS

Our approach to designing a sociotechnical audit can be applied more broadly. Here, we share insights for adapting our process to other contexts, such as auditing tenant screening algorithms.

7.1 How Do Audits Shift Power?

As researchers, we can develop many different kinds of audits, but all audits are not created equal [65, 80]. We can design audits that scrutinize systems of power and uplift impacted communities. In our work, we audited a sociotechnical system that concentrates power. Surveillance itself is an instrument of power, especially when used by police. Surveillance means to watch from above, from a position of power [19]. Our audit reverses the gaze and "watches from below" by evaluating police use of FRT surveillance. We built an external audit for regulatory bodies, policymakers, and civil society to challenge FRT and improve accountability to the public. We designed the audit questions to be answered based on publicly accessible information rather than information internal to the police. Thus, the audit can shift power from those using FRT to communities impacted by FRT.

Sociotechnical audits can expose the failures of systems with resources and influence. To promote greater accountability, researchers can design audits that enable the participation of parties independent from the auditee [30, 67, 80]. For example, researchers could design an external audit of tenant screening tools used by landlords in U.S. cities. This could reveal how transparent landlords are with the public and how communities who disproportionately face barriers to housing are impacted [83]. By rethinking audit tools, researchers can expand participation in algorithmic decision-making and center those most vulnerable to harm.

7.2 How to Make Audits Context and Jurisdiction-Specific?

Audits can localize what can at times seem like global or overreaching "algorthimic power". While technology often operates on a large scale, considering context requires coming down to a small scale. An important step when developing a sociotechnical audit involves understanding local injustices, social relations, and insights. This requires us to acknowledge specific impacted groups and forms of oppression [18].

Our audit is tailored to police use of FRT in England and Wales and is built on basic principles including fairness, transparency, and accountability. These principles could easily extend to another context such as tenant screening algorithms deployed in U.S. cities, but researchers would need to customize the audit to their specific context. As our work shows, there is no one-size-fits-all audit when ethical and legal considerations are included.

To build a sociotechnical audit that adapts our approach, researchers would need to consider the following aspects: community participation in the design of the audit, historical context of the sociotechnical system, oversight and governance structures of the auditee, and legal risks posed by the sociotechnical system.

Community participation: Researchers would need to engage with a variety of stakeholders in their specific context. As discussed in Section 4, we revised our audit based on feedback from community stakeholders focused on the use of policing technologies in the UK. More generally, researchers would need to actively involve local communities in their design process. For example, to develop an audit of tenant screening algorithms used in U.S. cities, researchers might engage with landlords and rental applicants, as well as researchers and policymakers focused on urban housing inequality. For community participation to be meaningful, researchers must work with stakeholders in the specific context and actually take into account their needs and recommendations [88].

Historical context: Researchers would need to consider historical power asymmetries in the particular context. In the UK policing system, there have been discriminatory policing practices that often target people of color and low-income communities. Adopting this audit more broadly would require researchers to understand how specific communities are disproportionately impacted by the algorithmic system. For example, in the U.S., Black, Hispanic, and Asian renters experience discrimination related to housing costs and quality [94]. Understanding such power dynamics can greatly inform how researchers design audit questions.

Governance structures: Researchers would need to identify the structures that govern or oversee the auditee. In our work, we learned that police forces in England and Wales operate independently and often have local ethics committees that can oversee the use of technology. We used this oversight structure in our audit by assessing whether local ethics committees provide independent oversight of FRT adoption. In the case of tenant screening algorithms, researchers might need to understand how the local government oversees landlords and housing issues such as racial disparities. Understanding such governance structures helps assess how the auditee is held responsible for its use of algorithms.

Legal risks: Finally, researchers would need to consider the legal system of the particular jurisdiction. Our audit was developed using specific legislation in England and Wales, such as the Data Protection Act 2018. For any sociotechnical audit, researchers must identify the legal rights with which the algorithmic system interferes. For example, to design an audit of tenant screening algorithms, researchers might consider protections from the Fair Housing Act of 1968 and state-level data privacy laws in the U.S. Researchers would then need to evaluate how the auditee, in this case, landlords, demonstrates compliance with the law. By considering these factors, researchers can adapt our approach and help improve accountability in how technologies are used in other contexts.

7.3 How Can Audits Reveal What Is Not Transparent?

One challenge when auditing is having access to necessary information. External audits, in particular, lack access to internal records and procedures of the auditee. While a lack of access can be a limitation, it can also be a finding, as it exposes what is hidden. For any sociotechnical system, we must not only ask what is transparent, but also what we are not able to see [5].

In our work, we were not able to access certain pieces of information from the two police forces that we audited. For example, even though the MPS conducted an internal evaluation of bias in their FRT software, they did not disclose the results. This makes it hard to assess the comprehensiveness of the evaluation. Additionally, neither the MPS nor SWP published the demographic breakdown for arrests resulting from the use of FRT. This makes it difficult to evaluate whether FRT perpetuates racial profiling. Such lack of transparency is not just a barrier; it is also an insight. It helps us understand how power is concentrated in the police and signals where there might be a need for change through advocacy or regulation.

Sociotechnical audits can help reveal what information is publicly unknown or inaccessible. For example, landlords may not disclose the reason for rejecting applicants when they make decisions using tenant screening algorithms. Audits can raise awareness of what is not transparent to the public. Such awareness can serve as a catalyst for action by motivating communities, regulators, decision-makers, and other stakeholders to call for change. Thus, audits can be useful even when information is not available. Next, we discuss the mechanisms by which audits can lead to change.

7.4 How Can Audits Lead To Accountability?

Audits can help hold developers, users, and other decision-makers accountable for the societal impacts of algorithms.

Accountability mechanisms: Audits can lead to accountability in several ways. First, the auditee may make changes to the

algorithmic system [28, 77, 102]. This may entail buy-in from the auditee during the auditing process. Second, an audit may lead to regulation or legal action [87]. This may require the participation of a body with oversight powers and legal authority over the auditee. Finally, the audit may spark public attention and demands for change [89]. This often involves researchers, journalists, or advocates disseminating the audit findings to the public.

We primarily focused on accountability through regulation and public attention. Three months after publishing our audit, an East London borough council passed a motion to ban police use of live FRT [48]. The councillor who proposed this motion referenced our audit, noting that the motion aligns with our findings of the MPS failing to meet ethical and legal standards [29]. Our results have also been shared publicly through news outlets such as *The Guardian* [37] and can motivate stakeholders in England and Wales to call for greater regulation, transparency, and accountability on FRT. For example, civil society can use our audit results as evidence in their own advocacy and litigation. More broadly, when designing a sociotechnical audit, researchers must consider how their audit will lead to accountability and can engage with stakeholders to strengthen the audit's impact.

Ongoing accountability: To help achieve continuing accountability, audits can be conducted periodically instead of just a single time. Recurring audits can be more effective at ensuring compliance [65, 80]. Ongoing assessment is especially important when evaluating a powerful institution that may continue unethical practices if unchecked. Additionally, the auditing process may need to be professionalized to ensure that evaluations are accurate and consistent over time [80].

We applied our sociotechnical audit to three cases, but in the future, researchers, journalists, civil society, or oversight bodies could conduct the audit on police use of FRT again. Our audit could be professionalized through the national data ethics governance body, which is underway by the UK government [54]. If this body has independence, resources, and legal authority, it could use our audit to help hold police accountable. When designing audits in other contexts, researchers can identify and work with entities that can conduct the audit repeatedly and demand change.

8 LIMITATIONS

While our audit reveals how police use of FRT perpetuates real-world harm, it nevertheless has some limitations. An important consideration is how police use of FRT compares to the baseline of police practice. For example, how does the accuracy or fairness of arrest change with FRT? To answer this question, we could conduct a field experiment and gather data on the arrest rates by police forces with and without FRT tools, which was outside the scope of this study. Nevertheless, our audit does demonstrate how FRT shifts power. Even if FRT is less discriminatory and generates fewer wrongful arrests, it can still be used to surveil marginalized groups.

At the same time, our audit is not exhaustive; it does not capture all harms related to police use of FRT. For example, FRT adoption can shift police suspicion and lead to over-policing unrelated to the technology [15, 44].

Additionally, using this audit to improve transparency alone cannot create accountability. This audit can expose harms in police use of FRT, but this is not equivalent to holding police accountable. However, transparency can be a starting point for accountability [5], as we discuss in Section 7.

Further, auditor independence is critical for this audit to provide meaningful scrutiny [80]. A police force auditing their own deployment of FRT would be similar to them marking their own homework [73]. However, even if the auditor is formally independent, they might have a conflict of interest with the police [67]. For instance, police councils and private companies hired by police may produce unreliable results.

While we encourage auditors to disclose key audit results, the degree of disclosure requires careful consideration as the results may be misused [30]. For example, police may engage in ethics washing where they exaggerate favorable findings. This can mask problematic practices and provide a false assurance of compliance with standards [47].

Finally, this audit's scoring system makes the audit simple but may miss complexities in an evaluation. The audit comprises yes/no questions that are scored with an explanation. Future work may entail designing "how" and "why" questions, or giving partial credit for answers and assigning weights to questions in order to prioritize critical ones.

9 CONCLUSION

Sociotechnical audits are useful for understanding and interrogating power asymmetries in society. By exposing the harms of how an algorithmic system is used, audits can motivate direct action toward greater accountability [65]. In this work, we designed a practical audit tool that scrutinizes a powerful sociotechnical system from the outside. The three police deployments of facial recognition that we examined all failed to meet ethical and legal standards for governing FRT, which emerged from our broad survey of *existing* laws, frameworks, and guidelines. The harms that we identified move beyond the concern of bias in facial recognition algorithms. In the cases we studied, the FRT deployments lacked independent oversight, transparent evaluations of discrimination, and evidence of a lawful interference with privacy rights. By revealing how the current use of FRT by police does not incorporate the known best practices for the safe and ethical use of AI systems, this work can strengthen calls for greater accountability and legislation.

We further provide insights to help researchers design their own sociotechnical audits. We discuss how such tools can be developed, how to adapt principles into practice, and how to bring a wide range of stakeholders into the design of an audit. Researchers can use our approach to empower affected communities to participate in crucial exercises of oversight and accountability. We hope this work contributes to the growth of tools that can examine and challenge how power is distributed in our digital society.

ACKNOWLEDGMENTS

Thank you to those who informed the work behind this audit through valuable feedback and discussions. These include: Andrew Strait, Ann Kristin Glenster, Areeq Chowdhury, Ben Bradford, Daragh Murray, Emmanuelle Andrews, Fraser Sampson, Griff Ferris, Ioannis Kouvakas, Jenny Brennan, Jun Pang, Jyoti Belur, Katrina Ffrench, Katy Watts, Lorna Woods, Madeleine Chang, Mher

Hakobyan, Nour Haidar, Nóra Ní Loideáin, Paul Quinton, Pete Fussey, and Tom McNeil. We also thank Angèle Christin for their helpful feedback on this paper and the Rotary Foundation for their support.

REFERENCES

[1] Ada Lovelace Institute and DataKind UK. 2020. Examining the Black Box: Tools for Assessing Algorithmic Systems, Identifying Common Language for Algorithm Audits and Impact Assessments. https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/.

[2] American Friends Service Committee. 2021. NEC Corp. https://investigate.afsc.org/company/nec.

[3] Amnesty International. 2018. Trapped in the Matrix: Secrecy, Stigma, and Bias in the Mat's Gangs Database. https://www.amnesty.org.uk/files/reports/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf.

[4] Amnesty International. 2022. USA: Facial recognition technology reinforcing racist stop-and-frisk policing in New York – new research. https://www.amnesty.org/en/latest/news/2022/02/usa-facial-recognition-technology-reinforcing-racist-stop-and-frisk-policing-in-new-york-new-research/.

[5] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media Soc.* 20, 3 (2018), 973–989.

[6] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. *Machine Bias.* Technical Report. ProPublica.

[7] Emily Apple. 2018. South Wales Police Under Fire for Using Facial Recognition Technology Against Protesters. https://www.thecanary.co/uk/2018/03/29/south-wales-police-under-fire-for-using-facial-recognition-technology-against-protesters/.

[8] Earl R. Babbie. 2020. *The Practice of Social Research.* Wadsworth Cengage Learning.

[9] Jack Bandy. 2021. Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 74 (apr 2021), 34 pages. https://doi.org/10.1145/3449148

[10] Pınar Barlas, Kyriakos Kyriakou, Styliani Kleanthous, and Jahna Otterbacher. 2019. Social B(eye)as: Human and Machine Descriptions of People Images. *Proceedings of the International AAAI Conference on Web and Social Media* 13, 01 (Jul. 2019), 583–591.

[11] BBC News. 2018. Black police leader says some forces 'still institutionally racist'. https://www.bbc.com/news/uk-england-42702432.

[12] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency.* 610–623.

[13] Ruha Benjamin. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code.* Polity, Medford, MA.

[14] Big Brother Watch. 2018. Face Off: The Lawless Growth of Facial Recognition in UK Policing. https://bigbrotherwatch.org.uk/wp-content/uploads/2018/05/Face-Off-final-digital-1.pdf.

[15] Big Brother Watch. 2020. Briefing on Facial Recognition Surveillance. https://bigbrotherwatch.org.uk/wp-content/uploads/2020/06/Big-Brother-Watch-briefing-on-Facial-recognition-surveillance-June-2020.pdf.

[16] Biometrics and Forensics Ethics Group. 2021. *Briefing Note on the Ethical Issues Arising from Public-Private Collaboration in the Use of Live Facial Recognition Technology.* Technical Report.

[17] Abeba Birhane. 2021. Algorithmic injustice: a relational ethics approach. *Patterns* 2, 2 (2021), 100205.

[18] Abeba Birhane, Elayne Ruane, Thomas Laurent, Matthew S. Brown, Johnathan Flowers, Anthony Ventresque, and Christopher L. Dancy. 2022. The Forgotten Margins of AI Ethics. In *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, June 21 - 24, 2022.* ACM, 948–958.

[19] Simone Browne. 2015. *Dark Matters: On the Surveillance of Blackness.* Duke University Press.

[20] Joy Buolamwini. 2022. *Facing the Coded Gaze with Evocative Audits and Algorithmic Audits.* Ph. D. Dissertation. Massachusetts Institute of Technology.

[21] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Conference on Fairness, Accountability and Transparency, FAT 2018, 23-24 February 2018, New York, NY, USA (Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, 77–91.

[22] Joy Buolamwini, Vicente Ordóñez, Jamie Morgenstern, and Erik Learned-Miller. 2020. Facial Recognition Technologies: A Primer. https://assets.website-files.com/5e027ca188c99e3515b404b7/5ed1002058516c11edc66a14_FRTsPrimerMay2020.pdf.

[23] Sheri Byrne-Haber. 2019. Disability and AI Bias. https://sheribyrnehaber.medium.com/disability-and-ai-bias-cced271bd533.

[24] Deborah Carr, Elizabeth Heger Boyle, Benjamin Cornwell, Shelley Correll, Robert Crosnoe, Jeremy Freese, and Mary C Waters. 2017. *Art and Science of Social Research.* WW Norton & Company.

[25] Central Digital and Data Office. 2018. Data Ethics Framework. https://www.gov.uk/government/publications/data-ethics-framework.

[26] Central Digital and Data Office and Office for AI. 2019. A Guide to Using Artificial Intelligence in the Public Sector. https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector.

[27] Albert Cherns. 1976. The principles of sociotechnical design. *Human relations* 29, 8 (1976), 783–792.

[28] Alexandra Chouldechova, Diana Benavides Prado, Oleksandr Fialko, and Rhema Vaithianathan. 2018. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Conference on Fairness, Accountability and Transparency, FAT 2018, 23-24 February 2018, New York, NY, USA (Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, 134–148.

[29] Areeq Chowdhury. 2023. https://twitter.com/AreeqChowdhury/status/1614926066476515330?s=20&t=j1wgV2DdCiD7of3lKHYL7A.

[30] Sasha Costanza-Chock, Inioluwa Deborah Raji, and Joy Buolamwini. 2022. Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. In *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, June 21 - 24, 2022.* ACM, 1571–1583.

[31] Amanda Coston, Neel Guha, Derek Ouyang, Lisa Lu, Alexandra Chouldechova, and Daniel E. Ho. 2021. Leveraging Administrative Data for Bias Audits: Assessing Disparate Coverage with Mobility Data for COVID-19 Policy. In *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021*, Madeleine Clare Elish, William Isaac, and Richard S. Zemel (Eds.). ACM, 173–184.

[32] Martin Coulter. 2020. A Black Man Spent 10 Days in Jail After He Was Misidentified by Facial Recognition, a New Lawsuit Says. https://www.businessinsider.com/black-man-facial-recognition-technology-crime-2020-12?r=MX&IR=T.

[33] Cradle. 2021. *Brick by Brick: How We Build a World Without Prisons.* Hajar.

[34] Angela Y Davis. 2011. *Are prisons obsolete?* Seven stories press.

[35] Digital Regulation Cooperation Forum. 2022. Auditing Algorithms: The Existing Landscape, Role of Regulators and Future Outlook. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1071554/DRCF_Algorithmic_audit.pdf.

[36] Vikram Dodd. 2017. Met police to use facial recognition software at Notting Hill carnival. https://www.theguardian.com/uk-news/2017/aug/05/met-police-facial-recognition-software-notting-hill-carnival.

[37] Vikram Dodd. 2022. This article is more than 3 months old UK police use of live facial recognition unlawful and unethical, report finds. https://www.theguardian.com/technology/2022/oct/27/live-facial-recognition-police-study-uk.

[38] Natalia Domagala. 2021. What is our new Algorithmic Transparency Standard? https://dataingovernment.blog.gov.uk/2021/11/29/what-is-our-new-algorithmic-transparency-standard/.

[39] Danielle Dwyer, Wesley Johnson, and Pa. 2010. Police apologise over CCTV in Muslim areas. https://www.independent.co.uk/news/uk/crime/police-apologise-over-cctv-in-muslim-areas-2094167.html.

[40] Laurel Eckhouse, Kristian Lum, Cynthia Conti-Cook, and Julie Ciccolini. 2019. Layers of bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior* 46, 2 (2019), 185–209.

[41] *R (Bridges) v. Chief Constable of South Wales Police.* 2020. Court of Appeal, Civil Division, case C1/2019/2670. https://www.judiciary.uk/wp-content/uploads/2020/08/R-Bridges-v-CC-South-Wales-ors-Judgment.pdf.

[42] *RMC and FJ v. Metropolitan Police Commissioner.* 2012. High Court, Queen's Bench Division, cases CO/12476/2010 and CO/5572/2011. https://www.judiciary.uk/wp-content/uploads/JCO/Documents/Judgments/r-rmc-fj-metropolitan-police-commissioner-22062012.pdf.

[43] Marc Faddoul, Guillaume Chaslot, and Hany Farid. 2020. *A longitudinal analysis of YouTube's promotion of conspiracy videos.* Technical Report.

[44] Pete Fussey, Bethan Davies, and Martin Innes. 2021. 'Assisted'facial recognition and the reinvention of suspicion and discretion in digital policing. *The British journal of criminology* 61, 2 (2021), 325–344.

[45] Peter Fussey and Daragh Murray. 2019. Independent report on the London Metropolitan Police Service's trial of live facial recognition technology. (2019).

[46] Clare Garvie, Alvaro Bedoya, and Jonathan Frankle. 2016. Perpetual Line-Up: Unregulated Police Face Recognition in America. *Georgetown Law Center on Privacy & Technology* 18 (2016).

[47] Ellen P Goodman and Julia Trehu. 2022. AI Audit Washing and Accountability. *Available at SSRN 4227350* (2022).

[48] Ruby Gregory. 2023. Councillor calls for ban on facial recognition cameras in East London borough. https://www.mylondon.news/news/east-london-news/councillor-calls-ban-facial-recognition-25990803.

[49] Patrick Grother, Mei Ngan, and Kayee Hanaoka. 2019. *Face recognition vendor test (fvrt): Part 3, demographic effects.* National Institute of Standards and Technology Gaithersburg, MD.

[50] Rebecca Heilweil. 2020. Big tech companies back away from selling facial recognition to police. That's progress. https://www.vox.com/recode/2020/6/10/21287194/amazon-microsoft-ibm-facial-recognition-moratorium-police.

[51] Kashmir Hill. 2020. Wrongfully Accused by an Algorithm. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

[52] Information Commissioner's Office. 2020. Guidance on the AI Auditing Framework: Draft Guidance for Consultation. https://ico.org.uk/media/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf.

[53] Tony Jefferson. 2012. Policing the riots: from Bristol and Brixton to Tottenham, via Toxteth, Handsworth, etc: Tony Jefferson tells the angry, ongoing story of rioting over the past 30 years. *Criminal justice matters* 87, 1 (2012), 8–9.

[54] Justice and Home Affairs Committee. 2022. *Technology Rules? The Advent of New Technologies in the Justice System.* Technical Report. House of Lords.

[55] Pratyusha Kalluri. 2020. Don't ask if artificial intelligence is good or fair, ask how it shifts power.

[56] Lauren Kirchner and Matthew Goldstein. 2020. Access Denied: Faulty Automated Background Checks Freeze Out Renters. https://themarkup.org/locked-out/2020/05/28/access-denied-faulty-automated-background-checks-freeze-out-renters.

[57] Ava Kofman and Ariana Tobin. 2019. Facebook Ads Can Still Discriminate Against Women and Older Workers, Despite a Civil Rights Settlement. https://www.propublica.org/article/facebook-ads-can-still-discriminate-against-women-and-older-workers-despite-a-civil-rights-settlement.

[58] PM Krafft, Meg Young, Michael Katell, Jennifer E Lee, Shankar Narayan, Micah Epstein, Dharma Dailey, Bernease Herman, Aaron Tam, Vivian Guetler, et al. 2021. An action-oriented AI policy toolkit for technology audits by community advocates and activists. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 772–781.

[59] Kyriakos Kyriakou, Pınar Barlas, Styliani Kleanthous, and Jahna Otterbacher. 2019. Fairness in Proprietary Image Tagging Algorithms: A Cross-Platform Audit on People Images. *Proceedings of the International AAAI Conference on Web and Social Media* 13, 01 (Jul. 2019), 313–322.

[60] Michelle S. Lam, Ayush Pandit, Colin Kalicki, Rachit Gupta, Poonam Sahoo, and Danaë Metaxa. 2023. Sociotechnical Audits: Broadening the Auditing Lens to Investigate Targeted Advertising. *Proc. ACM Hum.-Comput. Interact.* (2023).

[61] David Lammy. 2017. he Lammy Review Final Report: An Independent Review into the Treatment of, and Outcomes for, Black, Asian and Minority Ethnic Individuals in the Criminal Justice System. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/643001/lammy-review-final-report.pdf.

[62] Paul Lewis, Tim Newburn, Matthew Taylor, Catriona Mcgillivray, Aster Greenhill, Harold Frayman, and Rob Proctor. 2011. Reading the riots: investigating England's summer of disorder. (2011).

[63] Tamika Lewis, Seeta Peña Gangadharan, Mariella Saba, and Tawana Petty. 2018. *Digital Defense Playbook: Community Power Tools for Reclaiming Data.* Technical Report. Our Data Bodies.

[64] Liberty. 2022. Briefing on the Amended Surveillance Camera Code of Practice. https://www.libertyhumanrights.org.uk/wp-content/uploads/2022/01/Libertys-briefing-on-the-amended-Surveillance-Camera-Code-of-Practice-January-2022.pdf.

[65] Danaë Metaxa, Joon Sung Park, Ronald E. Robertson, Karrie Karahalios, Christo Wilson, Jeff T. Hancock, and Christian Sandvig. 2021. Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. *Found. Trends Hum. Comput. Interact.* 14, 4 (2021), 272–344.

[66] Jacob Metcalf, Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, and Madeleine Clare Elish. 2021. Algorithmic Impact Assessments and Accountability: The Co-Construction of Impacts. ACM, New York, NY, USA, 735–746.

[67] Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, Madeleine Clare Elish, and Jacob Metcalf. 2021. Assembling accountability: algorithmic impact assessment for the public interest. *Available at SSRN 3877437* (2021).

[68] Mozilla. 2021. Take control over your data with Rally, a novel privacy-first data sharing platform. https://blog.mozilla.org/en/mozilla/take-control-over-your-data-with-rally-a-novel-privacy-first-data-sharing-platform/.

[69] National Physical Laboratory and Metropolitan Police Service. 2020. Metropolitan Police Service Live Facial Recognition Trials. https://www.met.police.uk/SysSiteAssets/media/downloads/central/services/accessing-information/facial-recognition/met-evaluation-report.pdf.

[70] W. Lawrence Neuman. 2013. *Basics of Social Research: Qualitative and Quantitative Approaches.* Pearson Education.

[71] Safiya Umoja Noble. 2018. *Algorithms of Oppression.* New York University Press.

[72] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 6464 (2019), 447–453. https://doi.org/10.1126/science.aax2342

[73] 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).

[74] Privacy International. 2019. Submission to the Scottish Parliament's Justice Sub-Committee on Policing's Inquiry into Facial Recognition Policing. https://privacyinternational.org/sites/default/files/2019-11/19.11.01_JusticeSC_FRT_Evidence_PI_FINAL_2.pdf.

[75] Evani Radiya-Dixit. 2022. *A Sociotechnical Audit: Assessing Police Use of Facial Recognition.* Technical Report. Minderoo Centre for Technology and Democracy.

[76] Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 469–481.

[77] Inioluwa Deborah Raji and Joy Buolamwini. 2019. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2019, Honolulu, HI, USA, January 27-28, 2019*, Vincent Conitzer, Gillian K. Hadfield, and Shannon Vallor (Eds.). ACM, 429–435.

[78] Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. 2020. Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing. In *AIES '20: AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA, February 7-8, 2020*, Annette N. Markham, Julia Powles, Toby Walsh, and Anne L. Washington (Eds.). ACM, 145–151.

[79] 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*. ACM, New York, NY, USA, 33–44.

[80] Inioluwa Deborah Raji, Peggy Xu, Colleen Honigsberg, and Daniel E. Ho. 2022. Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance. In *AIES '22: AAAI/ACM Conference on AI, Ethics, and Society, Oxford, United Kingdom, May 19 - 21, 2021*, Vincent Conitzer, John Tasioulas, Matthias Scheutz, Ryan Calo, Martina Mara, and Annette Zimmermann (Eds.). ACM, 557–571.

[81] Divya Ramesh, Vaishnav Kameswaran, Ding Wang, and Nithya Sambasivan. 2022. How platform-user power relations shape algorithmic accountability: A case study of instant loan platforms and financially stressed users in India. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 1917–1928.

[82] Jenny Rees. 2020. Facial Recognition Use by South Wales Police Ruled Unlawful. https://www.bbc.co.uk/news/uk-wales-53734716.

[83] Eva Rosen, Philip ME Garboden, and Jennifer E Cossyleon. 2021. Racial discrimination in housing: how landlords use algorithms and home visits to screen tenants. *American Sociological Review* 86, 5 (2021), 787–822.

[84] Joe Ryan. 2021. *Reports of Misogyny and Sexual Harassment in the Metropolitan Police.* Technical Report. House of Commons.

[85] Javier Sánchez-Monedero, Lina Dencik, and Lilian Edwards. 2020. What does it mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 458–468.

[86] Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A Smith. 2019. The risk of racial bias in hate speech detection. In *Proceedings of the 57th annual meeting of the association for computational linguistics*. 1668–1678.

[87] Nathan Sheard. 2021. Banning Government Use of Face Recognition Technology: 2020 Year in Review. https://www.eff.org/deeplinks/2020/12/banning-government-use-face-recognition-technology-2020-year-review.

[88] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2022. Participation Is not a Design Fix for Machine Learning. In *Equity and Access in Algorithms, Mechanisms, and Optimization, EAAMO 2022, Arlington, VA, USA, October 6-9, 2022*. ACM, 1:1–1:6.

[89] Jacob Snow. 2018. Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots. https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28.

[90] South Wales Police. 2022. Facial Recognition App Pilot Results. https://www.south-wales.police.uk/news/south-wales/news/2022/ebr-apr/pilot-results-for-the-new-facial-recognition-app/.

[91] Chandler Nicholle Spinks. 2019. Contemporary housing discrimination: Facebook, targeted advertising, and the fair housing act. *Hous. L. Rev.* 57 (2019), 925.

[92] Latanya Sweeney. 2013. Discrimination in online ad delivery. *Commun. ACM* 56, 5 (2013), 44–54.

[93] TRUST San Diego Coalition. 2021. Surveillance Privacy Ordinances. https://sandiegotrust.org/20-Nov_Surveillance_Privacy_Ordinances.pdf.

[94] Margery Austin Turner, Robert Santos, Diane K Levy, Douglas A Wissoker, Claudia Aranda, and Rob Pitingolo. 2016. Housing discrimination against racial and ethnic minorities 2012: Executive summary. (2016).

[95] Frank Vanclay. 2003. International principles for social impact assessment. *Impact assessment and project appraisal* 21, 1 (2003), 5–12.

[96] Briana Vecchione, Karen Levy, and Solon Barocas. 2021. Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies. In *Equity and Access in Algorithms, Mechanisms, and Optimization.* Association for Computing Machinery, New York, NY, USA, Article 19, 9 pages.

[97] Alex S Vitale. 2021. *The end of policing.* Verso Books.

[98] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, et al. 2022. Taxonomy of risks posed by language models. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 214–229.

[99] West Midlands Police and Crime Commissioner. 2019. West Midlands Police's Ethics Committee: Terms of Reference. https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/07/Ethics-Committee-Terms-of-Reference-as-at-1-April-2019.pdf?x39505.

[100] White Coats for Black Lives. 2021. Racial Justice Report Card 2020-2021. https://whitecoats4blacklives.org/rjrc/.

[101] Patrick Williams. 2018. *Being Matrixed: The (Over)Policing of Gang Suspects in London.* Technical Report. StopWatch.

[102] Christo Wilson, Avijit Ghosh, Shan Jiang, Alan Mislove, Lewis Baker, Janelle Szary, Kelly Trindel, and Frida Polli. 2021. Building and Auditing Fair Algorithms: A Case Study in Candidate Screening. In *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021*, Madeleine Clare Elish, William Isaac, and Richard S. Zemel (Eds.). ACM, 666–677.