Responsible AI in Policing

Helena Webb
University of Nottingham
helena.webb@nottingham.ac.uk

Nicholas Fitzroy-Dale
Telemarq
nicholas@telemarq.com

Saamiya Aqeel
University of Nottingham
psysa11@exmail.nottingham.ac.uk

Anna-Maria Piskopani
University of Nottingham
anna-maria.piskopani@nottingham.ac.uk

Quentin Stafford-Fraser
Telemarq
quentin@telemarq.com

Christos Nikolaou
Telemarq
christos@telemarq.com

Liz Dowthwaite
University of Nottingham
liz.dowthwaite@nottingham.ac.uk

Derek Mcauley
University of Nottingham
derek.mcauley@nottingham.ac.uk

Christoper Hargreaves
University of Oxford
christopher.hargreaves@cs.ox.ac.uk

ABSTRACT

The deployment of AI-driven technologies in policing is often welcomed as an opportunity to enhance efficiency in dealing with crime. At the same time, however, these technologies pose risks around data bias, data protection, accuracy and privacy. In addition, socio-organisational factors present challenges to their deployment. In this short paper we provide an overview of the opportunities and risks associated with AI in policing, focusing on current developments in the UK. We discuss what is necessary for a responsible approach to deployment and highlight some of our own project work in this context. The *Trustworthy and Useful Tools for Mobile Phone Extraction* project shows that tools for the analysis of mobile phone data can include AI-driven features that are both useful and trustworthy. However, tools alone cannot address all the tensions and constraints that police work under. Therefore, an essential component of responsibility is to avoid overstating what AI can achieve.

CCS CONCEPTS

• **Applied Computing – Computer Forensics- Evidence collection, storage and analysis**;

KEYWORDS

Responsibility, Policing, Bias, Trust

ACM Reference Format:
Helena Webb, Nicholas Fitzroy-Dale, Saamiya Aqeel, Anna-Maria Piskopani, Quentin Stafford-Fraser, Christos Nikolaou, Liz Dowthwaite, Derek Mcauley, and Christoper Hargreaves. 2024. Responsible AI in Policing. In *Second International Symposium on Trustworthy Autonomous Systems (TAS '24), September 16–18, 2024, Austin, TX, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3686038.3686070

1 INTRODUCTION

Police work in the UK is carried out by 45 territorial police forces and 3 special police forces. Whilst these forces often conduct their activities very independently from each other (including in the procurement of new technologies [3]), overall direction is set by the National Police Chief's Council (NPCC). In 2021 the NPCC appointed its first Chief Scientific Adviser [23] and in 2024 its first lead for artificial intelligence (AI) [24]. These appointments demonstrate considerable enthusiasm for deploying novel technologies across forces. On taking up the role, the NPCC lead for AI said: ". . . AI presents opportunities for forces to test new ideas, be creative and seek innovative solutions to help boost productivity and be more effective in tackling crime" [24].

Policing is an exemplar of the opportunities and risks associated with the deployment of AI. In this short paper we overview some of the current uses of AI-driven technologies in policing in the UK in order to highlight these opportunities and risks. We reflect on what is needed for a responsible AI approach in this context and discuss our own relevant project work. The *Trustworthy and Useful Tools for Mobile Phone Extraction* project focuses specifically on the use of mobile phone data for the investigation of crime. AI-informed tools can assist with the extraction and analysis of mobile phone data; however, their usefulness and trustworthiness may be limited by technical and non-technical factors. We describe our development of an open-source tool, RIME (Responsible Investigation of Mobile Environments) to explore how AI features can be usefully and responsibly embedded in analytic tools. We also highlight the importance of not overstating the capacity for technological innovations alone to address entrenched tensions and constraints that complicate police work.

2 OPPORTUNITIES AND RISKS FOR AI IN POLICING

2.1 Opportunities

A prime motivation for bringing AI-driven technologies into policing is the promise of increased efficiency. Police forces across the UK have endured severe funding cuts, leading to claims that understaffed forces are losing capacity to deal with crime [12]. AI-driven technologies are often highlighted as a (partial) solution to this problem as they can conduct certain tasks with increased speed and

TAS '24, September 16–18, 2024, Austin, TX, USA

ACM ISBN 979-8-4007-0989-0/24/09
https://doi.org/10.1145/3686038.3686070

accuracy. This improves task efficiency and allows human working hours to be targeted more strategically.

One early area in which AI was adopted in this sector involves the use of machine learning for predictive policing. This has included efforts to identify hot spots where crimes are likely to occur, individuals who are likely to offend/reoffend, and individuals who might become victims of crime [4]. Advocates state that resources for crime prevention can then be directed towards the areas or individuals identified, and that this offers a more accurate and cost-effective approach than traditional policing methods [22].

Another area of interest lies in the automated analysis of images. One example is Automated Number Plate Recognition (ANPR) technologies, which capture a vehicle's number plate information when it passes a digital camera. The collected images are searched to identify vehicles of interest or to punish road traffic violations. The increased integration of AI techniques into ANPR has enhanced the accuracy and speed with which images can be analysed, including in changing and challenging environmental conditions [2]. A second example is automated facial recognition (AFR) technology. As the algorithms to support facial recognition have grown more sophisticated, they can be applied to an increasing range of image sources - such as CCTV cameras, Ring doorbells, mobile phones etc. – and can be used for both real-time and retrospective analysis. Captured images are cross-checked against the Police National Database (PND) or an individual police force's own database to identify people of interest. In September 2023, the UK government Minister for Policing urged for the greater use of AFR in criminal investigations [16].

2.2 Risks

Despite this enthusiasm, AI in policing is associated with various risks. Machine learning for predictive policing has raised familiar concerns over bias, with commentators noting that predictions produced by such systems will reproduce any biases in the data they are trained on [25]. In 2012 the Durham Police Force, in collaboration with the University of Cambridge, developed a decision support tool called HART (Harm Assessment Risk Tool). HART was designed to assist officers to determine whether to keep a suspect in custody or refer them to a crime diversion programme [8]. The tool was trained on data from previous cases and used 34 categories of information about a suspect to determine their risk of committing further offences, which in turn determined their eligibility for the diversion programme. The categories included postcode data; this risked causing a feedback loop in which the system drove police attention towards existing high crime areas and away from low crime areas [26]. Further criticisms of HART included its classification of offenders into 'crude groups' based on ethnic and socio-economic features, and its in-built avoidance of false negatives leading to an over-labelling of individuals as of at high risk of offending, meaning they would not be eligible for the diversion programme [6]. Durham Police stopped using HART in 2021, on resources grounds.

Even though the UK government has been advocating its use, automated facial recognition in policing has raised data protection concerns [18]. In 2019, the civil rights group Liberty supported a legal challenge against its use by South Wales Police. The practice was first ruled to be lawful but then unlawful on appeal. The appeal judgement [19] found that the use of AFR incurred proportionate interference with human rights but that the police force had not done enough to ensure the technology did not have racial or gender bias, and needed to do more to document who they were looking for and what evidence they had that people of interest would be in the monitored area. The force paused its use of AFR until a 2023 report published by the National Physical Laboratory [21] identified improved accuracy in facial recognition systems, including the capacity to mitigate for biases. Since then both the South Wales Police and Metropolitan Police have used AFR at public events such as concerts and sports matches. Discussions over the ethics and legality of AFR continue, with notes of caution highlighting the intrusiveness of taking and (temporarily) storing people's images and that, even if the technology is not biased, it will be disproportionately used to monitor certain demographics over others [5] [13].

Wider socio-organisational factors also present risks to the deployment of AI in policing. Despite policy commitments to technological innovation, police staff in the UK have expressed dissatisfaction and low confidence in ICT provision and training, with only 30% believing that their force invests wisely in technology [29]. Public opinion is also nuanced. Research suggests that people see a benefit to the use of facial recognition technology but fear the normalisation of surveillance and want appropriate safeguards to be in place [1]. The lack of transparency regarding algorithmic processes and decision-making is a barrier to confidence but can also lead to an over-trust in AI-based results [20]. Meanwhile, public confidence in the police overall is low [11], a consequence of high-profile scandals concerning misconduct and the mishandling of cases. The deployment of non-trusted technologies by a non-trusted sector can be seen as a considerable risk – in particular when the stakes are high since errors can cause miscarriages of justice.

3 AI IN POLICING: THE NEED FOR A RESPONSIBLE APPROACH

The brief overview above demonstrates that a responsible approach is needed to balance the opportunities and risks of deploying AI in policing. A responsible approach involves taking a future-facing interest in identifying the intended and unintended impacts of innovation, plus attending to societal concerns to ensure the acceptability of innovation [27]. Current evidence suggests that some unintended negative impacts of AI in policing have been identified but not necessarily mitigated. In addition, public perspectives and values do not fully align with current deployments. Vallor's 2023 declaration [32] provides a useful means to conceptualise the attitudes that need to be adopted to foster responsible AI in policing, and to assess the degree to which shifts towards these attitudes have been made so far. Firstly, Vallor states that responsibility needs to be *accepted*. The NPCC can be seen to be taking steps to acknowledge its responsibilities through the publication of a covenant for using AI [24] that includes ethical principles and mechanisms for governance. It can also be seen to understand responsibility as *relational* by emphasising the importance of transparency to secure public trust and confidence [24]. The efforts to address potential biases in facial recognition technologies can be seen to *attend to vulnerability* but concerns remain over data protection and privacy intrusion as well as biases and feedback loops in predictive policing systems, Similarly, *sustainability* remains an issue, as seen in the ceased deployment of interventions such as HART and known problems around the retention of staff in policing [14].

4 *TRUSTWORTHY AND USEFUL TOOLS FOR MOBILE PHONE EXTRACTION:* EXPLORING RESPONSIBLE AI IN POLICING

Our ongoing project *Trustworthy and Useful Tools for Mobile Phone Extraction* investigates digital tools in a particular area of policing. When police are investigating a crime, they may request (or compel) a suspect, witness or complainant to hand over their mobile phone. Data from the phone is collected and analysed. This process, known as mobile phone extraction (MPE), can identify important evidence to resolve cases. However, since modern phones typically hold very large amounts of data, processes of analysis can be slow and efficient [30]. Failures to properly analyse data have been associated with the collapse of court cases, e.g. [10]. Digital tools to assist with analysis do exist but can be inaccessible due to cost and/or technical complexity. Additionally, collecting data can pose privacy risks. Until recently, the default was for all data on a phone to be extracted; this meant that large amounts of personal data might become visible to the police, defence and others across the criminal justice system, even if it is not relevant to the case. As highlighted by civil rights organisations [7] and the Information Commissioner's Office [17], phone owners are often fearful of non-relevant data being seen and potentially used to undermine them. Even with new guidance [15] stating that police should conduct a more selective extraction, especially when the phone owner is a witness or complainant, concerns remain over the handling of third-party data [31].

These problems have affected public confidence in MPE, in particular in the conduct of cases where phone owners may be vulnerable [9]. Our project explores the capacity for trustworthy digital tools to address the problems by making MPE practices more efficient and more respectful of privacy. Our project is guided by a responsible approach and involves a series of activities [28] including the development of a mobile phone extraction tool, RIME (Responsible Investigation of Mobile Environments). Complying with the new guidance, RIME attends to data privacy; it is designed to expose a subset of the contents of a phone for investigation, rather than a complete capture of all data on the device. The data subset can be filtered by criteria such as time and contact names for targeted, and efficient, analysis. The continuing development of RIME has allowed us to gauge the extent to which AI can help to address the problems associated with MPE. In the remainder of this section, we illustrate this with two examples.

4.1 Understanding limits to the potential for AI-driven features in MPE tools

Data availability limits the kinds of AI features that can be embedded in MPE tools. Ethical and legal considerations inhibit the collection of large personal mobile datasets for research purposes and synthetic data may lack quality. Consequently, it can be difficult to develop AI features that can make robust inferences since they are not trained on suitable data.

For example, we developed a pie chart feature for RIME, which categorises messages within a mobile phone dataset according to criteria such as 'potential argumentative nature' and 'of evidentiary interest.' This is intended to capture references to potential illegal activities and other terminology that might be useful when investigating a case. A machine learning model, a CNN text classifier, categorised the messages. However, when it came to training this model, the data used was problematic. The model, when trained, will pick up on positives based on its training dataset, but finding a dataset that captures unique nuances in language and naturalistic terminology is difficult. In the context of 'evidentiary interest,' people rarely mention crimes explicitly in messages, instead using slang or emojis. Slang can vary significantly by region, and favouring one region over another can introduce undesirable bias. Dataset creation is therefore highly complex and curating such a dataset takes time and research to reach suitable levels of specificity.

Despite the caution noted above, AI can offer significant benefits when it is used in a way that avoids the risks of implicit bias. In the development of RIME we have prioritised using well-tested AI techniques ahead of the most recent innovations. We have also chosen traditional, non-AI, techniques where they are shown to be quicker and/or more accurate. An example is RIME's pseudonymisation feature. When looking at a data subset, a user can select for real names and phone numbers to be replaced with autogenerated, but indexed, alternatives. If/when the individual concerned becomes a person of interest in the case, the pseudonymisation can be removed. This is a very useful feature to protect third-party privacy (which *attends to vulnerability*) and is partially AI-driven. RIME uses a Named Entity Recognition (NER) model to identify names, and traditional techniques to identify phone numbers and email addresses. RIME maps identified Personal Identifiable Information (PII) to a token, so the same anonymised token is re-used for the same PII. This works well for phone numbers and emails as they can always be canonicalised but is a challenge for names (e.g. 'Bob', 'B', and 'Robert' may all refer to the same person). We experimented with using large language models (LLM) to identify PII, but NER gave better results for the sizes of LLM that could be run locally.

4.2 The potential impact of AI-driven tools is contingent on the context in which they are used

Our stakeholder engagement work has revealed socio-organisational dynamics that influence how MPE practices are conducted. For instance, even when the police prefer a selective extraction, the defence and/or the Crown Prosecution Service (CPS) may seek a full one [28]. This can lead to a full extraction being undertaken regardless of the existence of tools to support selective extraction. As another example, police staff can find the outputs of MPE tools very difficult to work with. Tools produce a report that captures the extracted data file formats such as pdf, excel or .ufd. These can be analysed manually or within the tools themselves. Given the amounts of data typically involved, these files can be extremely large. Consequently, police staff reviewing/analysing phone data as part of the investigation of cases find can them very difficult to interact with – particularly given that they may be working on relatively old and slow computers. When talking to stakeholders, a common request we hear is for RIME data reports to be in a more accessible size and format. So, we are aware there is a need to balance this kind request alongside exploration of state-of-the-art AI techniques.

Another important dynamic is the lack of trust around MPE. This is something we attend to in the development of RIME, prioritising openness and replicability. A dataset collected within RIME can be made available to third parties for analysis using their own tools. This allows the prosecution and defence, or even an independent authority, to conduct their own analysis of a dataset. In addition, RIME is an open-source tool, with the code publicly available on the University of Nottingham Horizon Digital Economy Github account. This is another transparency and trustworthiness measure since the source code is open to inspection. We also hope that it will be a measure for *sustainability* whereby interested community members support its ongoing development and add new features. This is particularly important as mobile phone technologies are always changing; it is necessary to ongoingly extend RIME's plugins to read data stored in previously unsupported formats or data generated by the latest mobile apps.

5 CONCLUSION

The deployment of AI in policing presents opportunities and risks. These relate to the features of AI-driven technologies themselves and to wider socio-organisational dynamics. A responsible approach to AI in policing requires attention to both these factors. Our project work highlights that the inclusion of AI in tools for mobile phone extraction needs to be carefully researched and scoped. It is important to consider the context in which tools will be used in order to maximise their usefulness and trustworthiness. Whilst AI undoubtedly offers benefits, it cannot alone address all the problematic issues associated with contemporary policing. It is important for a responsible approach to acknowledge this and not over-state the promise of AI.

ACKNOWLEDGMENTS

We are grateful to all the project participants who have contributed to this study. The Trustworthy and Useful Tools for Mobile Phone Extraction project is funded by a pump prime award from the UKRI Trustworthy Autonomous Systems Hub (EP/V00784X/1). The initial development of RIME was funded by another UKRI Trustworthy Autonomous Systems Hub award, Digital Forensics Platform. Due to the ethical restrictions, supporting data in the research study cannot be made available.

REFERENCES

[1] Ada Lovelace Institute. 2019. Beyond face value: public attitudes to facial recognition technology.

[2] Muhammad Babelle Ahmad, Umar Farouk Musa, Muntaka Dahiru, and Mustapha Babatunde Abimbola. 2024. Advantages of Automated License Plate Recognition Technology. Eng. Technol, 4(1), pp.10-15

[3] Julie Ayling 2012. A good buy: promoting probity in police procurement. In Handbook of Policing, Ethics and Professional Standards (pp. 90-101). Routledge.

[4] Alexander Babuta. 2017. Big data and policing: an assessment of law enforcement requirements, expectations and priorities. Royal United Services Institute for Defence and Security Studies.

[5] bbc.co.uk. 2023. Facial recognition tech: Liberty 'police racism' claim. Retrieved 12 May, 2024 from https://www.bbc.co.uk/news/uk-wales-65214494

[6] Big Brother Watch. 2018. A closer look at Experian big data and artificial intelligence in Durham police. April 2018 Retrieved 10 April, 2024 from https://bigbrotherwatch.org.uk/blog/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

[7] Big Brother Watch. 2019. Digital Strip Searches: the police's data investigations of victims. July 2019. Retrieved 10 April, 2024 from https://bigbrotherwatch.org.uk/wp-content/uploads/2019/07/Digital-Strip-Searches-Final.pdf

[8] Centre for Public Impact. 2018. Durham Constabulary's AI decision aid for custody officers. Retrieved May 2, 2024 from https://www.centreforpublicimpact.org/assets/documents/ai-case-study-criminal-justice.pdf

[9] Centre for Women's Justice. 2020. Stop the 'digital strip search' of rape victims like me. March 2020. Retrieved Jan 10, 2024 from https://www.centreforwomensjustice.org.uk/new-blog-1/2020/3/13/stop-digital-strip-search

[10] CPS. 2018. Joint review of the disclosure process in the case of R v Allan. Jan 2018. Retrieved Dec 18, 2023 from https://www.cps.gov.uk/publication/joint-review-disclosure-process-case-r-v-allan

[11] Vikram Dodd. 2024. Only 40% of people in England trust their police force, research reveals. theguardian.com. April 18, 2024. Retrieved May 1, 2024 from https://www.theguardian.com/uk-news/2024/apr/18/only-40-of-people-in-england-trust-their-police-force-research-reveals

[12] Elisa Facchetti, E., 2021. Police infrastructure, police performance, and crime: Evidence from austerity cuts. Job Market Paper, Queen Mary University of London, 4

[13] European Data Protection Board. 2023. Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement. Version 2.0 Adopted 26 April 2023.

[14] Matthew Fright, Nick Davies and Gil Richards. 2023. Retention in public services. Institute for Government. October 2023.

[15] Home Office. 2023. Extraction of information from electronic devices: code of practice. Retrieved Jan 14, 2024 from https://www.gov.uk/government/consultations/extraction-of-information-from-electronic-devices-code-of-practice/extraction-of-information-from-electronic-devices-code-of-practice-accessible

[16] Home Office. 2023. Speech. Policing Minister: Police Superintendent's Association Conference. September 2023. Retrieved 18 March, 2024 from https://www.gov.uk/government/speeches/policing-minister-police-superintendents-association-conference

[17] Information Commissioner's Office. 2020. Mobile Phone data extraction by police forces in England and Wales. June 2020. Retrieved Jan 11, 2023 from https://ico.org.uk/media/about-the-ico/documents/2617838/ico-report-on-mpe-in-england-and-wales-v1_1.pdf

[18] Information Commissioner's Office. 2023. ICO statement in response to parliamentarian's letter on facial recognition technology. Dec 2023. Retrieved May 31, 2024 from https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2023/12/ico-statement-in-response-to-parliamentarians-letter-on-facial-recognition-technology/

[19] judiciary.uk. 2020. Press Summary. The Queen V The Chief Constable of South Wales Police and others [2020]EWCA Civ 1058. Retrieved March 18, 2024 from https://www.judiciary.uk/wp-content/uploads/2020/08/R-Bridges-v-CC-South-Wales-ors-Press-Summary.pdf

[20] Lauren Leffer. 2024. Too much trust in AI poses unexpected threats to the scientific process. Scientific American. March 18, 2024. Retrieved 10 April, 2024 from https://www.scientificamerican.com/article/trust-ai-science-risks/

[21] Tony Mansfield. 2023. Facial Recognition technology in law enforcement equitability study final report. National Physical Laboratory.

[22] M.R. McGuire. 2021. The laughing policebot: automation and the end of policing. Policing and society, 31(1), pp.20-36.

[23] NPCC. Office of the Chief Scientific Adviser. Retrieved May 20, 2024 from https://www.npcc.police.uk/our-work/office-of-the-chief-scientific-adviser

[24] NPCC. NPCC Welcomes first lead for artificial intelligence. Retrieved May 20, 2024 from https://news.npcc.police.uk/releases/npcc-welcomes-first-ever-lead-for-artificial-intelligence-ai

[25] Marion Oswald and Alexander Babuta. 2019. Data analytics and algorithmic bias in policing. Royal United Services Institute for Defence and Security Studies.

[26] 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), pp.223-250

[27] Richard Owen, Phil Macnaghten and Jack Stilgoe. 2020. Responsible research and innovation: From science in society to science for society, with society. In Emerging Technologies (pp. 117-126). Routledge.

[28] Anna-Maria Piskopani, Helena Webb, Christopher Hargraves, Liz Dowthwaite, Nicholas FitzRoy-Dale, Quentin Stafford-Fraser, Christos Nikolaou and Derek McAuley. 2024. Trustworthy and Useful Tools for Mobile Phone Extraction, Proceedings of the ETHICOMP 2024 21st International Conference on the Ethics and Social Impacts of ICT March 2024, pp. 256-258.

[29] Policinginsight.com. 2018. Police investigations being hindered by ‘outdated’ IT systems, officers and staff warn. Retrieved 11 May, 2024 from https://policinginsight.com/news/copacc-launches-latest-police-ict-user-report-frontline-insight-from-48-uk-forces/

[30] Eric Priezkalns. 2022. UK police failing to gather evidence from huge backlog of mobile phones. commsrisk.com. 8 Dec, 2022. Retrieved May 1, 2024 from https://commsrisk.com/uk-police-failing-to-gather-evidence-from-huge-backlog-of-mobile-phones/

[31] Privacyinternational.org. 2021. Policing Bill: An unsatisfactory debut on the statute books for mobile phone extraction. June 2021. Retrieved Jan 10, 2024 from https://privacyinternational.org/news-analysis/4586/policing-bill-unsatisfactory-debut-statute-books-mobile-phone-extraction

[32] Shannon Vallor. 2023. Edinburgh Declaration on responsibility for responsible AI. medium.com. Retrieved February 11. 2023 from https://medium.com/@svallor_10030/edinburgh-declaration-on-responsibility-for-responsible-ai-1a98ed2e328b