Article

Police Journal: Theory, Practice and Principles

The Police Journal: Theory, Practice and Principles
2025, Vol. 0(0) 1–21

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0032258X251372357
journals.sagepub.com/home/pjx
Sage

RUDI: An evidence-based police-centric guide for approaching the development of algorithmic models in policing

Hazel Sayer[1], **Tamara Polajnar**[2] and **Ruth Spence**[3]

Abstract

Police are increasingly adopting or considering machine learning algorithms (MLAs) to enhance their processes, appealing for their ability to process large volumes of information and predictive capabilities. Lack of national guidance on developing and implementing algorithms means police forge ahead in an exploratory manner. To address this, we developed a practical police-centric framework and guide: RUDI (Rationale, Unification, Development, Implementation), a framework designed to mitigate concerns raised regarding the use of algorithms in policing. This report outlines our work with two police forces to develop RUDI, highlighting the challenges of algorithmic policing and how RUDI can mitigate these concerns.

Keywords

machine learning, algorithms, framework, transparency, policing

Introduction

Since 2010, United Kingdom (UK) government austerity measures have seen unprecedented cuts to police funding, which has significantly affected policing in the UK. There

[1]Bournemouth University, Dorset, UK.
[2]Independent Consultant, Cambridgeshire, UK.
[3]Middlesex University, London, UK.

Corresponding author:
Hazel Sayer, Bournemouth University, Fern Barrow, Poole, Dorset BH12 5BB, UK.
Email: hazel@herethical.ai

has also been a surge in violent crime since 2013, and while not conclusive, this has partly been attributed to falling police numbers and resources (Draca and Langella, 2020). These budget reductions have necessitated a more strategic allocation of resources and a re-evaluation of decision-making processes within the criminal justice system, which has coincided with the rise of increased technological capabilities (Ferguson, 2017). Therefore, policing is increasingly turning to the use of artificial intelligence (AI) and data analytics in a broader shift to 'data-driven policing' (Kearns and Muir, 2019). Algorithmic policing is an umbrella term which covers the use of data analytics and machine learning algorithms (MLAs) by police to assist them in certain aspects of policing, such as crime prediction, resource allocation, identifying targets for intervention and guiding police decision-making, especially with deployment of its personnel and resources (Chan and Bennett Moses, 2016; Perry, 2013). Both data analytics and MLAs involve analysing large amounts of data to identify patterns, data analytics does this through statistical analysis whilst MLAs predict unknown outcomes based on past results using features in the data (Police-ML, 2024). Police believe this shift towards a data-driven approach can help improve efficiency (Schlehahn et al., 2015). This is reflected in increased central funding for 'digital transformation', as well as encouragement for police to increase their digital solution capabilities, incentivising police to pursue algorithmic policing (The Law Society of England and Wales, 2019).

While there have been several attempts to improve the transparency of the use of algorithms in public services, most notably the Algorithmic Transparency Recording Standard (ATRS, GOV.UK, 2023) and Information Commissioners Office AI auditing framework (Information Commissioners Office, 2020), they lack the practical guidelines needed to help develop MLAs. We introduce a framework, specifically aimed at policing, which is contextualised on the operational knowledge of the UK policing practices and the technical understanding of development and implementation of effective MLAs.

Background

There are several concerns with the introduction of algorithms in policing, including the suitability of policing data for such analysis (Davies, 2023), as well as a lack of national guidelines on how police should proceed. This has led to the development of frameworks to guide forces on raising standards in developing algorithms. One of the most comprehensive frameworks is ALGO-CARE (a mnemonic for Advisory, Lawful, Granularity, Ownership, Challengeable, Accuracy, Responsible, Explainable; Oswald et al., 2018). ALGO-CARE aims to ensure algorithmic big data technologies can be used for policing purposes without violating data protection legislation (Babuta, 2017; Bland, 2020). Additionally, as part of the government's National Data Strategy, the Central Digital and Data Office and the Centre for Data Ethics and Innovation are helping public sector organisations to be more transparent about their use of algorithmic tools by recording decisions in the ATRS (GOV.UK, 2023).

Yet despite the emergence of these frameworks, a significant challenge persists in translating and implementing such knowledge into effective policing strategies (Nichols et al., 2019). Academia faces the task of rendering complex research into accessible 'police speak' (Stanko, 2007), underscoring the existing divide between academic insights and practical police applications (Goode and Lumsden, 2018). Furthermore, there is the challenge of bringing together data science and domain expertise, each frequently work in isolation and find it difficult to find common ground (Babuta, 2017; Mao et al., 2019). A prevailing misconception exacerbates the issue: falsely assuming data scientists, internal or external to the force, are responsible for developing data science capabilities, when in fact, there needs to be a collaborative effort (Viaene, 2013). To ensure their products are relevant and fit for purpose data scientists cannot work in silos without police input. This may be particularly difficult when presently there are no national guidelines that set out the practicalities involved in developing MLA capabilities. Indeed, without fully knowing what is involved, police may embark on MLA projects not realising they do not have the required expertise to develop and implement the modelling techniques required, or that simpler approaches to data analysis could be used. Missing key stages of development, such as a robust approach to tackling bias, can exacerbate rather than ameliorate problems, not to mention waste police resources if projects are abandoned or found to be ineffective. Therefore, there is currently a need for a framework that actively outlines the practical steps police need to take to develop, build and maintain applications that depend on MLAs.

Lack of evidence regarding benefits

One crucial and over-arching criticism of algorithms in policing is that often the benefits remain unclear. This is partly because there are so few independent evaluations of algorithms and frequently, police do not release the results of any internal testing procedures (Shapiro, 2017). This means it is difficult to assess the accuracy or effectiveness of algorithms (EUCPN, 2022). Where there is external scrutiny, the results can be mixed. For example, 'COMPAS', risk assessment software was used in the US courts to assess criminal defendants' likelihood of recidivism. Its developer claimed COMPAS was able to accurately predict recidivism for any offence, offences against persons and felony offences over a 5-year period (Brennan et al., 2009). However, non-profit organisation ProPublica used the same benchmarks as the creators of the algorithm and demonstrated COMPAS was remarkably ineffective in predicting criminal behaviour (Angwin et al., 2022). Similarly, Durham Constabulary's HART model, which predicted mid-risk subjects' future offending for an out-of-court disposal intervention was evaluated and only found to be marginally better at predicting mid-risk subjects than custody sergeants (Urwin, 2017).

Partly this may be because algorithms cannot consider the individual context officers would typically factor into decision-making (Bland, 2020), instead generalising from overarching patterns to individual behaviours. This also shifts focus from the societal and individual circumstances causing violent and/or criminal behaviour to whether there are enough data to pre-empt it (Andrejevic, 2017). This shift in focus to algorithmic policing is considered a 'distraction' from dealing with the root causes of crime (Verma, 2022).

Police resourcing crisis driving the perceived need for algorithms

Nevertheless, police are under pressure in terms of resource allocation and greater need for efficiency and there is a concern the increase in digital transformation funding incentivises police to develop digital technology as opposed to finding alternative solutions (The Law Society, 2019). Indeed, the premise for developing and implementing algorithms may be misguided. Police believe algorithms will target resources more effectively and save money in the long term, seemingly offering a low-cost solution to the lack of resources in UK police forces (Babuta and Oswald, 2020). However, the need for additional resources, such as specialist staff for ML model development and maintenance, and additional personnel to manage data quality and respond effectively to algorithmic outputs, is frequently underestimated (McKay and Richard, 2022). Certainly, Durham and Kent Constabularies both ceased using algorithms due to issues with resourcing and costs (Durham Constabulary, 2021; Nilsson, 2018).

Bias

There are also more practical concerns with algorithms. The data can over-represent some groups and areas, particularly socially deprived areas (Richardson et al., 2019). There is an abundance of research concerning algorithms reproducing, reinforcing and magnifying racial, socioeconomic, demographic and gender biases (e.g., Berk and Hyatt, 2015; Brantingham 2017; Harcourt, 2015; Huq, 2018; Starr, 2014). For example, until it was decommissioned in February 2024, the Gangs Matrix was a predictive algorithm used by The Metropolitan Police that scored subjects on their propensity to commit serious organised or gang related crimes. It was called into question after being found to largely target young, black men (Dodd, 2020). Similarly, PredPol was more likely to target low-income, black communities compared to affluent, white communities with similar rates of drug crimes in the United States (Lum and Isaac, 2016). ProPublica found COMPAS incorrectly considered black defendants to be twice as likely to commit crimes as white defendants (Angwin et al., 2022). Durham's HART model included predictors such as age, gender, occupation, family composition and postcode, as well as the offender's history of criminal behaviour, the inclusion of such predictors could be discriminatory and perpetuate indirect biases towards areas marked by deprivation (Palmiotto, 2021).

Further, there is also concern about potential bias in how police interact with algorithms. For instance, automation bias is the propensity to rely on an algorithmic output rather than one's own expertise. This may become particularly problematic if police do not understand the reasons for the proposed decision, because it can induce compliance as algorithms are often presented as 'outperforming' human expertise (Hildebrandt, 2018). Therefore, over time police may become deskilled and ever more reliant on algorithms (Babuta and Oswald, 2019; Hildebrandt, 2017). There is also concern around confirmation bias, the tendency to look for information that confirms or strengthens beliefs and disregard information that is incongruent. Officers have been found to be susceptible to this, with a tendency to ignore AI generated recommendations if they disagree with their professional judgement, even if the reasons for the AI result are explained (Selten et al., 2022).

Poor data quality

Relatedly, algorithms are often purported to be 'objective' (Kearns and Muir, 2019). However, this ignores the numerous decisions and subjective choices fed into the algorithm, including the data driving these algorithmic models which are often quite limited, and insufficiently representative of the population and crime committed. Police records can be considered to reflect enforcement rather than the true crime rate, and to some extent measure an interaction between crime, policing strategies and community-police relations (Lum and Isaac, 2016; Robinson and Koepke, 2016). This means algorithms will focus on some crimes and perpetrators whilst missing others for which there is limited or no data. For example, a high proportion of the most serious domestic violence perpetrators have no prior record for domestic violence, which means any algorithm based only on known abuse could not identify such cases (Bland, 2020). Further, algorithmic policing draws on limited in-force data. For example, the data used for Durham Constabulary's HART model did not include data from other police forces, or national databases such as the Police National Database, thus offering only a snapshot of an individual's past criminality. Currently, including this is a logistical impossibility, given police data is so fragmented and disparate across different systems. Further, amalgamating this data elicits concerns about what is necessary and proportionate, and it is argued such use of data contravenes Article 8 of the European Convention on Human Rights – the right to respect for privacy (Liberty, 2019). Having limited in-force data to draw on is therefore one reason why the model can only *inform* police decision making, not replace it, as officers have access to other resources to give them a bigger picture (Bland, 2020).

Additionally, research has highlighted those entering the data, often frontline police officers, tend to prioritise investigative tasks over precise data entry (Terpstra and Kort, 2017). Crucial information is often missing or incomplete, lacks sufficient detail, has not been updated as the case progresses, or contains inaccuracies such as spelling errors (Burcher and Whelan, 2018; Dencik et al., 2018; O'Connor et al., 2022; Sanders and Henderson, 2013). Where attempts are made to standardise data entry, individuals are still required to subjectively categorise incidents, which can lead to inconsistencies (Sanders and Henderson, 2013). Poor data quality has implications for the subsequent data analyses and creates decisions for analysts, for instance should cases with missing values be deleted or the missing values imputed? Any decision affects the outcome, and potentially any algorithm's accuracy and reliability.

Transparency, explainability, and accountability

Given these data issues, it is troubling that there is also a lack of transparency and accountability in how data is used to inform policing decisions, as well as an inherent difficulty in the interpretation of how the data-points used by MLAs contribute to their output. This means the process by which police algorithms arrive at their output has been described as 'incomprehensible' (Schlehahn et al., 2015) and 'inscrutable' (Mittelstadt et al., 2016), and means we often cannot understand, even in hindsight, why an algorithm has made certain decisions (EUCPN, 2022). This lack of transparency has seen growing concern regarding the use of predictive algorithms amongst the public and civil liberty groups alike (Robinson, 2019), raising fears these systems are inaccessible for scrutiny, making them impossible to evaluate or legally contest, particularly in the context of evidence disclosure (Grace, 2019). The opportunity to challenge the outcome of a decision made by an algorithm, the underlying reason that decision was made or the integrity of the data that was used to inform the algorithm is practically non-existent (Grzymek and Puntschuh, 2019; Hildebrandt, 2015). Nevertheless, Article 6 of the European Court of Human Rights stipulates a person's right to know the reasons for decisions which adversely and significantly affect that individual, meaning this lack of transparency poses a risk to the right to a fair trial and due process (European Court of Human Rights, 2022a). This was highlighted when the Information Commissioner's Office in the UK took an important step in issuing an enforcement notice against the Metropolitan Police in relation to 'lack of transparency' and 'inaccuracy' of its Gangs Matrix (Grace, 2019).

Guidelines for disclosing information about the use of data analytics by public bodies is also lacking, and Freedom of Information requests by Dencik et al. (2018) regarding the use of data analytics in local authorities were successful less than 25% of the time, with refusal being cited as interfering with police activity or being against the commercial interests of a private company. This means important decision-making factors remain opaque. An example of this would be the decision to privilege one type of error over another when developing the model. This involves a choice between minimising the number of subjects falsely flagged (a false positive) against those who are missed by the algorithm (a false negative). Organisations must weigh up the possible consequences of false positives (privacy and intrusion, waste of resources) and false negatives (missing those who go on to commit serious offences), including reputational damage to the organisation.

Nevertheless, despite these challenges police are already experimenting with algorithmic decision support systems without legal and regulatory frameworks in place (Kearns and Muir, 2019). Indeed, recent systematic reviews show that police forces globally are increasingly applying quantitative decision-support tools, including ranking algorithms such as PROMETHEE and other multi-criteria approaches, to aid in patrol planning, resource allocation, and hotspot identification (Costa and Silva, 2024). However, these implementations often exhibit methodological inconsistencies, indicating a need for greater care and structure in the development stage, particularly in contexts where oversight is limited and outcomes may carry significant ethical implications. Therefore, from a pragmatic point of view, there is an urgent need for ensuring they do so with guidance based on best practice recommendations to bridge the gap between the academic evidence base, front line policing and data science. This article sets out the development of a police-centric guide designed to work towards (1) mitigating the concerns raised regarding MLAs, (2) bridging the gap between data science and policing by facilitating a shared understanding when developing MLAs, and (3) fostering transparency to improve reproducibility and explicability of MLAs.

Methodology

We adopted a two-stage approach, firstly in developing the framework and the secondly on testing the framework. The first stage was conducted in collaboration with a large English police force (police force A) who had already built a pilot algorithm for prioritising suspects. Their pilot model's primary function was to identify and rank potential suspects based on various criteria, including risk factors and historical data points. Its secondary function was to estimate suspect's future risk. We drew on the experiences and knowledge that they had gleaned over the period of developing this pilot model, complemented by a narrative literature review, incorporating best practice findings as well as practical considerations relevant to algorithmic development and deployment in policing contexts.

Interviews with police force A were conducted between June and September 2023. A purposive, selective sampling strategy was used, whereby participants were identified on the basis of their involvement with developing the model and could provide detailed information about the unification of data and modelling process. Four single participant and two joint interviews (with two and three participants) were completed. The interviews were recorded on MS Teams. The semi-structured interview questions were designed to focus on the data preparation, development of the pilot model and the intended operationalisation of the model. To draw out their processes, the challenges encountered during the development of the pilot model, the integration of the model into existing policing workflows, and the perceived benefits and limitations of the algorithm. Thematic analysis was conducted by the first author, using an inductive approach whereby themes were developed from the data to identify recurring ideas. These codes were then grouped into broader thematic categories that informed the development of the framework.

The second stage focused on testing the framework's Conceptualisation and Rationale sections with a smaller force (police force B) who wanted to introduce algorithmic modelling to identify and rank high harm suspects. This was to ensure the framework developed was applicable to algorithms beyond that developed by the initial force, and to test the framework's adaptability, ease of use and effectiveness when applied to a different police force's context. Testing included sharing the framework with force B and providing guidance on setting up a multi-disciplinary team. The research and multi-disciplinary teams from force B then met eight times between April and July 2024 to discuss and develop a business case. During these meetings the research team provided advice to help the team develop their business case utilising the RUDI framework and solicited feedback on how the framework could be improved. Additionally, four one-to-one interviews and three small discussion groups were conducted over the same timeframe, with a total of 12 officers of varying rank and key police stakeholders. These were recorded on MS Teams. A purposive, selective sampling strategy was used, whereby participants were identified from the specialist unit investigating domestic abuse, for which the model would have intended use, or chosen because their role had previously been, or would be impacted by algorithms. Both interviews and focus groups utilised a semi-structured format, focussing on the identification of what type of model would be the best fit for the needs of the force, what data was held that would need to be fed into the model, how it would fit into operational procedures, and who would be responsible for each stage of the development, implementation and maintenance of such a model. Once again, the first author conducted thematic analysis on both interview and focus group data using an inductive approach and resulting themes, including barriers and feedback from force B were fed back into the development of the framework. The process for the two stages of developing the framework and testing it are explained in Figure 1.

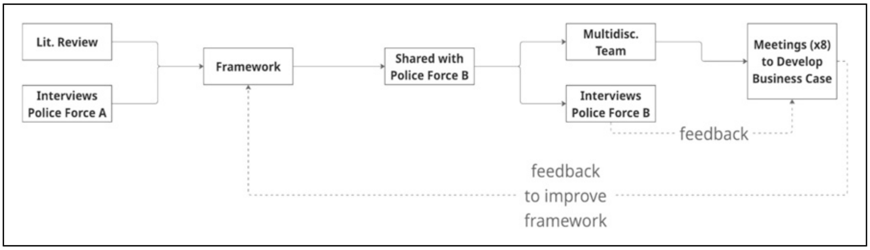

Figure 1. Methodological process of the development of the RUDI framework.

Results and discussion

The RUDI framework draws on insights from three stages of the research project: a narrative literature review, the experiences and lessons learned by Force A during the development of their pilot algorithm, and the reflections of Force B as they defined the problem they sought to address and built a business case for algorithmic modelling. RUDI places the key stages identified during the research project's development and testing phases into a process that police can follow to develop MLAs whilst mitigating some of the associated issues. This framework is presented first, then the themes identified during the development and testing phases and their implications are discussed using the RUDI framework.

Final framework – RUDI

RUDI was developed to be specific to complex models that use ML techniques, the outcome of which are likely to drive police action and have possible consequences for members of the public (e.g., models that prioritise people or predict future criminal behaviour or locations). Existing frameworks such as ALGO-CARE have been developed as guidance to ensure that police algorithms are ethical, lawful are responsibly deployed. RUDI is complementary to existing frameworks by covering the practical steps a force should take, from conceptualising a problem that can be aided by the use of MLAs, right to implementation and maintenance, and who should be involved and accountable at each stage. In other words, RUDI focuses on *how* to develop, build and maintain MLAs, limiting the subjectivity in which different forces can go about developing MLAs, and by focusing on having the right team and expertise to carry out each stage. To this end, RUDI covers four main stages:

Rationale: Outlines the stakeholders required to establish a multi-disciplinary team that will collaborate on the rationale for MLAs. Provides a template with pertinent questions designed to create a business case which documents the algorithmic journey, makes decisions explicit and justifies actions during unification, development and implementation.

Unification: Outlines a process for merging data sources together for modelling, quality checking data and ensuring its validity and reliability for use in modelling.

Development: Provides a guide to the steps involved in building and testing models, evaluating bias, performance and limitations, and choosing the preferred model given these factors. Additionally, it includes data and model cards for documenting these processes.

Implementation: Suggests ways of testing how the model will feed into current practice and how to maintain the model over time.

RUDI highlights that project evolution is not a linear process (see Figure 2). However, conceptualisation of the problem should come first, with the final step being implementation of the proposed solution. Rationale, unification, and development all pose reasonable entry points into project execution, and throughout the project lifecycle, the team will have to revisit each of these stages until an acceptable solution is found for the problem outlined at the conceptualisation stage. Implementation includes model monitoring and updates which will necessitate a return to earlier stages as well. There is no 'right way' to develop data models; it involves a series of decisions that must be made whilst balancing multiple competing concerns. RUDI provides recommendations based on best practice, offers guidance that sets out some of the decisions and issues involved, and provides templates for forces so they can document the process and improve transparency, justifiability, and accountability (see Appendix A). It *does not* prescribe an exact formula for algorithmic modelling.

The RUDI framework serves as a practical toolkit for any police force considering the deployment of machine learning algorithms (MLAs). It is freely available as a

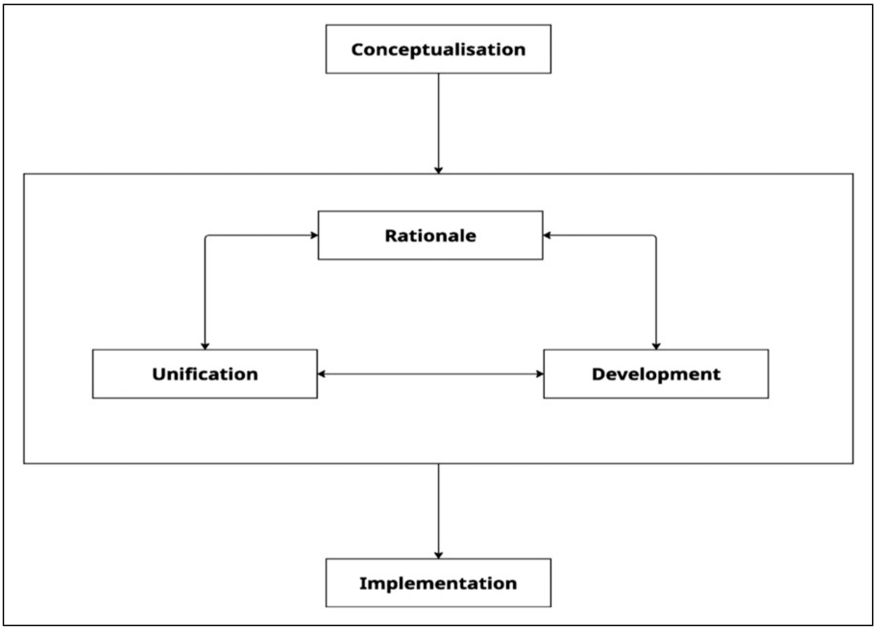

Figure 2. The RUDI process.

downloadable PDF via the ‘Resources’ section of www.police-ml.com, alongside supporting templates designed to assist with the ethical and transparent development of MLAs. The website provides step-by-step guidance throughout the algorithmic lifecycle, from initial concept to deployment and maintenance. Future plans include the development of training materials to help police forces maximise the value of RUDI. An AI-powered chatbot is also available on the site to respond to common questions and offer guidance on responsible algorithm design.

Rationale

Multi-disciplinary team oversight. The first step in developing the rationale for why a particular MLA solution is required is bringing together the right team of people to define and oversee the project. Indeed, the role of experts in the development of algorithms is extremely important and the consequences of not having the right people in place on a team can encourage misuse of the model and cause reputational risks for the police (Bland, 2020; Sayer et al., 2024). Assigning responsibility via a multidisciplinary team at the earliest opportunity ensures consistent accountability for the model’s performance at all stages and allows internal auditing of the process to take placethat is, those in charge of developing the model are different from those reviewing the model’s performance**.** Certainly, responsibility to rigorously test not only the model but also the governance implications such as fairness, bias, and transparency, falls on the team collectively (Busuioc, 2021). Babuta (2017) notes that when police forces seek to develop a model, they often work completely independently from model developers; likely to rarely or never meet. However, collaboration between data scientists and domain experts ensures a shared understanding of diverse business concepts, preventing potential misinterpretations and misrepresentations that could lead to failure (Viaene, 2013). Modelling that adopts a ‘conversation mode’ allows data scientists to tap into the intimate knowledge held by domain experts regarding business ideas and their context. This approach facilitates the identification and challenge of any underlying assumptions or biases in the data. Through constructive and focused conversations, a comprehensive, holistic view of the business context can be attained.

Develop a business case. Once established, the main task of the multi-disciplinary team is to develop a business case. This is because the introduction of modelling represents a substantial commitment, involving both time and resources. Therefore, at the outset, it is important to establish a foundation for modelling through properly conceptualising the project’s aims and developing a shared understanding of key definitions, for example, ‘high harm’, who the initiative will target, and what mechanism will be best for doing this. Additionally, it ensures the modelling is necessary and that both the benefits and long-term costs are considered (Babuta and Oswald, 2020). Without such forethought there is a high chance the deployment of a model will fail (McKay and Richard, 2022).

A clear business case facilitates a shared understanding of how the model works, as inability to fully explain the algorithmic output can lead to a lack of transparency (Mittelstadt et al., 2016), directly affecting accountability for the model and its decisions (Yeung, 2018). Additionally, it allows for the entire process to be auditable, enabling transparency on all decisions, which helps create an institutional memory and documentation so if key personnel leave the project, the learning does not leave with them. It can also be used to establish specific, measurable and achievable evaluation criteria at the outset of the project. Currently it is difficult to ascertain whether algorithmic models achieve what they set out to do, that is, reduce crime – particularly because they are often data-driven rather than theory-driven (Saunders et al., 2016). Flawed or insufficient theoretical underpinning can result in a range of issues and unintended consequences, such as distortions in the analysis that would likely produce patterns that turn out to be spurious correlations (Kaufmann et al., 2019). However, documenting decisions and processes readies the model for transparent validation, evaluation and the comparative studies needed if an evidence base is to be created regarding the costs and benefits of AI applications in policing (EUCPN, 2022). These can then be kept in mind during the development phase, as well as used for more formal evaluations, adding to the evidence-base. It should be noted that projects can take an 'exploratory' approach, which involves starting with the data analysis and not establishing a clear purpose for analysis until after the insights have been generated if a business case is then clearly established (Babuta and Oswald, 2020).

A business case also allows forces to think through likely costs and capacity issues. Certainly, there are concerns that police are using MLAs as a solution to bypass the real issue – lack of personnel and resources (Babuta and Oswald, 2020). Therefore, algorithmic tools as opposed to non-technological solutions may be unnecessary and incur many potential unrealised costs when money might be better spent elsewhere. For example, most forces' storage infrastructures are unlikely to be suitable for real-time dynamic data and data mining applications and cloud technology can be costly, especially on an ongoing basis (Robinson, 2019). Equally, there are staff costs associated with upskilling officers to lead on data projects, as well as potentially teaching staff how to develop and maintain algorithmic models that depend on ML techniques. Additionally, extra staff or resources may be required to ensure there is capacity to properly respond to the algorithm's output, as well as allow for external evaluations of the model and training of end-users on how to use the model.

Moreover, documenting the aims and suitable and unsuitable use cases ahead of time guards against 'function creep'. This can also be done through purpose limitation at the development stage (Fantin et al., 2020). Function creep occurs when the use of an algorithm expands beyond its original purpose, resulting in applications that are incongruent with its original purpose, which can create unintended and uncontrolled consequences. For example, a study by Saunders et al. (2016), evaluated the Strategic Subjects List (SSL), which was intended to identify high risk individuals deemed to be suitable for bespoke intervention to reduce the prevalence of gun violence. Instead, they found being identified by the SSL was related to an increase in the likelihood of being arrested for a shooting, suggesting the algorithm drove greater surveillance and enforcement rather than increasing other preventative strategies.

Unification

Data quality. The poor data quality of police forces has been recognised (Office for Statistics Regulation, 2024). One of the most straightforward ways to mitigate this is to improve force data recording and quality, ensuring robust oversight and regular data quality assessments (EUCPN, 2022). MLA performance is dependent on the data that it is being trained and tested on. This concerns the quality of input data, which can be mediated programmatically to a degree (Jain et al., 2020), but still affects the predictive outcomes through missingness (Davies et al., 2022). Also, to make accurate predictions MLAs require access to all the knowledge that might be pertinent to decision making. For example, if we are predicting the likelihood of an escalation into serious crime, we need to know of all the previous relevant instances of criminal patterns of escalation and whether they were interrupted by successful interventions, prison, removal from the area, or death. This relevant data is currently spread across disparate databases and, therefore, requires unification.

Development. Before the development of an MLA solution takes place, thought must be given both to what data is pertinent, why it is required and its quality, as well as if there is the necessary expertise within force to develop an algorithm.

Data needs. To fulfil the intended purpose of an algorithm and help to ensure transparency and auditability, Felzmann et al. (2020) advocate a transparency-by-design approach, requiring discussion and agreement of what data specifically is needed. Gebru et al. (2021) proposed using standard documentary procedures for every 'component' of the model, no matter how simple or complex, in similar fashion to the electronics industry. Forces are encouraged to document their decisions not only in the business case template, but also using data and model cards. These allow potential costs to be considered at the start of the project and make issues like data quality clear. The templates and cards can be completed iteratively throughout the project to help ensure the model and the process of its development is explainable and transparent. The *data cards* document an overview of the properties of the dataset used in the model, whilst the *model cards* document processes so that models can be audited, compared and evaluated. These prompt data scientists to describe any data quality issues and the resulting model limitations, which can feedback into data quality improvement efforts. Additionally, the cards can explain how the model was developed, for instance how data was split into training and validation sets or decisions on cut offs, such as what values would be considered errors for example, a suspect aged 5, so it is clear to internal and external stakeholders. These cards improve model transparency, explicability, and reproducibility. The documentation could also be used to make algorithms more public facing, enabling external scrutiny and alternatives to be tested, especially regarding categories that can serve as proxies for protected characteristics like race and sex (Palmiotto, 2021).

Nevertheless, to some extent ML algorithms are going to be a black box - an explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible. In these circumstances, justification of what data is going into the model and why and linking it to theory (Oswald et al., 2018), as well as other explicability measures (e.g. traceability, auditability and transparent communication on system capabilities) can help, provided the system respects fundamental rights.

Data science needs. Currently only a few UK forces have a dedicated data science (DS) team. This leaves them at a disadvantage when it comes to navigating the current data and technology landscape both from an organisational perspective and while pursuing criminal behaviour in a world where crime is increasingly data and machine learning powered. In-house DS teams drive data quality and optimisation of resources, while also providing a ready source of expertise for evaluation of external vendors or understanding of digital crime. As a result, many forces are considering setting up their own teams, but there are several considerations.

Police forces are understandably security conscious when it comes to adopting new technologies. However, this can negatively affect internal data science teams who need access to latest programming tools and specialised computer hardware. It is important that efforts are made to provide adequate computational power on a secure network, so that the team is not limited in their creativity when examining potential solutions.

Likewise, ML skills are in high demand and require extra specialisation and experience. Given the speed of advancements in AI, which can already do much of the preparatory work such as merging datasets and querying databases, future competencies will likely focus on specialist technical skills, the ability to oversee ML tools and integrating data insights into wider decision-making processes (Afzal and Panagiotopoulos, 2024; Muir and O'Connell, 2025). However, hiring and integration into an existing organisation can be difficult. Funding is still scarce, which makes it difficult for public services to invest in salaries that are competitive with the wider market and can also be perceived unfair in comparison with other employees. Alternative solutions can include upskilling current employees with interest and aptitude or setting up regional data science centres that operate as service points for a few forces but are more tightly integrated than a national body would be. It is important that the teams include a depth of scientific experience as well as administrative staff that is versed in project management and communication with stakeholders from frontline to executive levels.

Mitigating bias in development and implementation. During the development process itself, it is crucial to check for and mitigate against bias. Certainly, one of the biggest obstacles in achieving success with algorithmic policing is unrepresentative data which can result in 'algorithmic bias', which can creep in during any stage of data modelling projects (Barocas and Selbst, 2016). This is where algorithms systematically and unfairly discriminate against certain individuals or groups of individuals in favour of others (Friedman and Nissenbaum, 1996), as shown when the previously discussed COMPAS algorithm disproportionally flagged African Americans. This is likely due to biased data starting with the fact African Americans are more likely to be arrested and incarcerated in the US due to historical racism and inequalities in the criminal justice system. This would then be reflected in the training data and used to make suggestions about whether a defendant should be detained. This demonstrates that if historical biases are factored into the model, it will make the same kinds of incorrect judgments people do (Lee et al., 2019). Models which create direct or indirect discrimination on the grounds of protected characteristics contravene the Equality Act (UK, 2010) and Article 14 of European Court of Human Rights (2022b). Although it may be impossible to eliminate all biases from police data, every effort should be made to do so (Oswald et al., 2018) and model development should assess, document and outline the steps taken to reduce bias where possible.

Police have statutory obligations to assess all information that pertains to their case and their judgement and expertise takes precedence. Algorithmic policing should be used in an advisory context and as one factor alongside the many other sources that police are statutorily obliged to consider (Bland, 2020; Oswald et al., 2018). By ensuring the algorithm is used in an advisory context only and keeping the human decision-maker in the loop, risks to individuals by means of false positives and false negatives can be mitigated (Oswald et al., 2018). Moreover, training end-users in how to interpret the algorithmic outputs ensures they are aware of police dataset limitations, recognise the incomplete picture they present, and understanding that an algorithm cannot consider perpetrators or crimes for which no data is on the system. The EUCPN (2022) recommends specific training on algorithm limitations and individual/organisational responsibilities. Exposure to failures during training can also help guard against complacency, whereas relying on telling users about the limitations, and warning them to always verify does not sufficiently reduce automation bias (Skitka et al., 2000). Establishing protocols for end-users to challenge algorithmic decisions can empower them to have the confidence to do so.

Over-reliance on algorithmic outputs can leave users unable to understand, explain and justify their decisions and make full use of algorithm-provided information (Klein et al., 2006). Automation bias can become more problematic over time, particularly if police become accustomed to working with algorithmic outputs and view them as highly accurate and reliable (Alon-Barkat and Busuioc, 2022; Prinzel et al., 2001). As a growing number of domain experts begin to use algorithmic outputs when making decisions, training can facilitate them in understanding, overseeing, explaining and justifying algorithmic decision-making, which is fundamental to mitigating negative social impacts or harms (Simkute et al., 2021). Similarly, confirmation bias, where police officers form a hypothesis and interpret information to prove that hypothesis, can be a problem (Selten et al., 2022). Further to assessing model accuracy, evaluations could also consider possible proxy effect on officer deskilling, for example, decision-making atrophy and over reliance on algorithms to the detriment of their own judgement (Bland, 2020).

Implementation

Model maintenance. Developing and building models in-house rather than commissioning external agencies or paying for software has three major advantages. Firstly, the model will be specific to the aims outlined in the business case and the data of each force; secondly, any model needs to be maintained over time and thirdly, having access to the training data enables biases specific to the that dataset to be properly assessed and mitigated (Babuta and Oswald, 2019). Understandably, many forces may not have the available personnel or capacity to have an internal data science team. However, using an external agency to fulfil some of these roles may lead to problems when maintaining the model over time, and using software that has been developed externally, which is subject to licence fees can prove to be very costly. For example, Durham Constabulary's HART model was developed in collaboration with statistical experts based at the University of Cambridge (Oswald et al., 2018). However, the force did not then have the resources required to constantly refine and refresh the HART model on an ongoing basis and as a result they stopped using it (Durham Constabulary, 2021).

Data properties can change over time for a variety of reasons, for example, fluctuations in crime, changes in investigative practices, changes in input software. This can lead to model performance changing over time and without proper test procedures in place errors can creep in undetected. Likewise, technology is continuously improving, and better performance might be possible, and not using the best available technology for key functions can be considered unethical.

* An important caveat to note is that with the advancement of MLAs, particularly generative models, there are new and evolving challenges - notably adversarial attacks such as prompt injection, model inversion, and data poisoning, which are well-documented in computational research. The RUDI framework does not offer direct technical prescriptions on cybersecurity, to avoid offering potentially outdated advice in a rapidly changing domain overseen by specialist bodies such as the National Cyber Crime Unit and the Police Digital Service, we recognise the importance of aligning policing practices with these emerging risks.

Conclusion

The growing integration of machine learning (ML) algorithms in policing brings with it both significant potential for improving efficiency and fairness, and critical ethical, transparency and implementational challenges that must be addressed. To bridge the gap between the academic evidence base and everyday policing, the RUDI framework stands as a comprehensive and practical police-led guide that supports law enforcement in navigating the utilisation of complex ML algorithms. By following the steps outlined in RUDI, police can establish clear rationale and protocols for the development, implementation, and oversight of ML models, thereby fostering a culture of transparency within policing and with the public. The framework's attention to explicability helps to promote clearer communication between police and data scientists, as well as a deeper understanding of what data feeds into the model and its output, as well as support efforts to mitigate algorithmic bias, which has been a persistent issue in predictive policing systems and other law enforcement applications.

Moreover, RUDI's focus on data quality and model reproducibility equips police forces with the tools to critically assess the datasets they use, helps enable the comparison of different models and provides clear guidelines for model evaluation. The framework contributes to the standardisation of algorithmic practices, making them more comparable and verifiable across different jurisdictions. This not only enhances the accountability of policing agencies but also contributes to the broader academic and professional discourse on evidence-based policing.

Whilst RUDI cannot and does not overcome all issues inherent with police modelling, by embedding transparency, systematic evaluation, and continuous improvement into the development of ML algorithms, RUDI offers a promising approach for minimising the risks and ensuring that law enforcement agencies adopt more responsible, transparent, and evidence-based practices in their use of technology. The RUDI framework represents a critical step forward in bridging the gap between academic research and real-world policing.

ORCID iDs

Hazel Sayer https://orcid.org/0000-0003-2521-7903
Tamara Polajnar https://orcid.org/0009-0005-0875-2746
Ruth Spence https://orcid.org/0000-0002-6197-9975

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Office of the Police Chief Scientific Advisor.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

References

Afzal M and Panagiotopoulos P (2024) Data in policing: an integrative review. *International Journal of Public Administration* 48: 411–430.

Alon-Barkat S and Busuioc M (2022) Human-AI interactions in public sector decision making "Automation Bias" and "Selective Adherence" to algorithmic advice. *Journal of Public Administration Research and Theory* 33: 153–169.

Andrejevic M (2017) To pre-empt a thief. *International Journal of Communication* 8: 879–896. Available from: https://ijoc.org/index.php/ijoc/article/view/6308 (accessed 13 July 2024).

Angwin J, Larson J, Mattu S, et al. (2022) *Machine Bias. Ethics of Data and Analytics*. Auerbach Publications, pp. 254–264.

Babuta A (2017) Big data and policing. *An Assessment of Law Enforcement Requirements, Expectations and Priorities*. Available from: https://www.rusi.org/explore-our-research/publications/occasional-papers/big-data-and-policing-assessment-law-enforcement-requirements-expectations-and-priorities (accessed 9 January 2024).

Babuta A and Oswald M (2019) Data analytics and algorithmic bias in policing. *RUSI*. Available from: https://www.rusi.org/explore-our-research/publications/briefing-papers/data-analytics-and-algorithmic-bias-policing (accessed 10 February 2024).

Babuta A and Oswald M (2020) Data analytics and algorithms in policing in England and Wales: towards a new policy framework. *RUSI*. Available from: https://www.rusi.org/explore-our-research/publications/occasional-papers/data-analytics-and-algorithms-policing-england-and-wales-towards-new-policy-framework (accessed 10 February 2024).

Barocas S and Selbst AD (2016) *Big Ddata's disparate impact*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. Available from: https://papers.ssrn.com/abstract=2477899 (accessed 15 December 2023).

Berk R and Hyatt J (2015) Machine learning forecasts of risk to inform sentencing decisions. *Federal Sentencing Reporter* 27(4): 222–228.

Bland M (2020) Algorithms can predict domestic abuse, but should we let them? In: Jahankhani H, Akhgar B, Cochrane P and Dastbaz M (eds). *Policing in the Era of.* AI and Smart Societies (Advanced Sciences and Technologies for Security ApplicationsSpringer, 139–155.

Brantingham PJ (2017) The logic of data bias and its impact on place-based predictive policing. *Ohio State Journal of Criminal Law* 15: 473. Available at: https://heinonline.org/HOL/LandingPage?handle=hein.journals/osjcl15anddiv=32andid=andpage= (accessed 12 January 2024).

Brennan T, Dieterich W and Ehret B (2009) Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior* 36(1): 21–40.

Burcher M and Whelan C (2018) Social network analysis as a tool for criminal intelligence: understanding its potential from the perspectives of intelligence analysts. *Trends in Organized Crime* 21(3): 278.

Busuioc M (2021) Accountable artificial intelligence: holding algorithms to account. *Public Administration Review* 81(5): 825–836.

Chan J and Bennett Moses L (2016) Is big data challenging criminology? *Theoretical Criminology* 20(1): 21–39.

Costa J and Silva M (2024) Multicriteria decision-making in public security: a systematic review. *Mathematics* 12(11): 1754.

Davies K (2023) Data-driven policing. In: Davies GM, Beech AR and Colloff MF (eds) *Forensic Psychology: Crime, Justice, Law, Interventions*. 4th edition. Wiley-Blackwell.

Davies K, Spence R, Cummings E, et al. (2022) Understanding sexual violence and factors related to police outcomes. *Frontiers in Psychology* 13: 977318.

Dencik L, Hintz A, Redden J, et al. (2018) Data scores as governance: investigating uses of citizen scoring in public services project report. Available from: https://eprints.goldsmiths.ac.uk/id/eprint/37291/1/data-scores-as-governance-project-report2.pdf (accessed 12 January 2024).

Dodd V (2020) Met removes hundreds from gangs matrix after breaking data laws. *The Guardian*. Available from: https://www.theguardian.com/uk-news/2020/feb/15/met-removes-hundreds-from-gangs-matrix-after-breaking-data-laws (accessed 5 January 2024).

Draca M and Langella M (2020) *Law, Order and Austerity: Police Numbers and Crime in the 2010s*. Cage Research Centre. Available from: https://warwick.ac.uk/fac/soc/economics/research/centres/cage/news/03-06-20-advantage_magazine__summer/article-3/3._law_order_and_austerity.pdf (accessed 7 January 2024).

Durham Constabulary (2021) AI can predict reoffending, University Study finds. Available from: https://www.durham.police.uk/News/News-Articles/2022/January/AI-can-predict-reoffending-university-study-finds.aspx (accessed 4 August 2023).

Equality Act (2010) *GOV.UK*. Available from: https://www.legislation.gov.uk/ukpga/2010/15/contents (accessed 10 January 2024).

EUCPN: European Crime Prevention Network (2022) *Artificial Intelligence and Predictive Policing: Risks and Challenges*. EUCPN. Available from: https://eucpn.org/sites/default/files/document/files/PP%282%29.pdf (accesssed 30 November 2023).

European Court of Human Rights (2022a) Guide on Article 6 of the European Convention on Human Rights: right to a fair trial. Available from: https://www.echr.coe.int/documents/d/echr/guide_Art_6_eng (accessed January 2024).

European Court of Human Rights (2022b) Guide on Article 14 of the European Convention on Human Rights and on Article 1 of protocol no. 12 to the convention. Available from: https://www.echr.coe.int/documents/d/echr/Guide_Art_14_Art_1_Protocol_12_ENG (accessed January 2024).

Fantin S, Emanuilov I, Vogiatzoglou P, et al. (2020) Purpose limitation by design as a counter to function creep and system insecurity in police artificial intelligence *UNICRI special collection on AI in criminal justice*. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3679850 (accessed 12 December 2023).

Felzmann H, Fosch-Villaronga E, Lutz C, et al. (2020) Towards transparency by design for artificial intelligence. *Science and Engineering Ethics* 26(6): 3333–3361.

Ferguson AG (2017) *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York University Press.

Friedman B and Nissenbaum H (1996) Bias in computer systems. *ACM Transactions on Information Systems* 14(3): 330–347.

Gebru T, Morgenstern J, Vecchione B, et al. (2021) Datasheets for datasets. *Communications of the ACM* 64(12): 86–92.

Goode J and Lumsden K (2018) The McDonaldisation of police-academic partnerships: organisational and cultural barriers encountered in moving from research on police to research with police. *Policing and Society* 28(1): 75–89.

GOV.UK (2023) Algorithmic transparency recording standard hub. Available from: https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub (accessed 7 January 2024).

Grace J (2019) Machine learning technologies and human rights in criminal justice contexts. Available from SSRN 3487454. DOI: 10.2139/ssrn.3487454.

Grzymek V and Puntschuh M (2019) What Europe knows and thinks about algorithms results of a representative survey. *Bertelsmann Stiftung Eupinions*. Available from: https://aei.pitt.edu/102582/ (accessed 19 January 2024).

Harcourt BE (2015) Risk as a proxy for race. *Federal Sentencing Reporter* 27(4): 237–243.

Hildebrandt M (2015) *Smart Technologies and the End (s) of Law: Novel Entanglements of Law and Technology*. Edward Elgar Publishing.

Hildebrandt M (2017) Law as computation in the era of artificial legal intelligence. Speaking law to the power of statistics. *University of Toronto Law Journal* 68(suppl. 1): 12–35.

Hildebrandt M (2018) Algorithmic regulation and the rule of law. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 376(2128): 20170355.

Huq AZ (2018) Racial equity in algorithmic criminal justice. *Duke Law Journal* 68: 1043.

Information Commissioners Office (2020) *Guidance on the AI Auditing Framework*. ICO. Available from: https://ico.org.uk/media/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (accessed 9 January 2024).

Jain A, Patel H, Nagalapatti L, et al. (2020) Overview and importance of data quality for machine learning tasks. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '20)*. Association for Computing Machinery, pp. 3561–3562.

Kaufmann M, Egbert S and Leese M (2019) Predictive policing and the politics of patterns. *The British Journal of Criminology,* 59(3).

Kearns I and Muir R (2019) Data driven policing and public value. *The Police Foundation*. Available from: https://www.police-foundation.org.uk/wp-content/uploads/2010/10/data_driven_policing_final.pdf (accessed 17 January 2024).

Klein G, Moon B and Hoffman RR (2006) Making sense of sensemaking 2: a macrocognitive model. *IEEE Intelligent Systems* 21(5): 88–92.

Lee NT, Resnick P and Barton G (2019) *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*. Brookings Institute, Vol. 2. Available from: https://policycommons.net/artifacts/4141276/algorithmic-bias-detection-and-mitigation/4949849/ (accessed 2 February 2024).

Liberty (2019) *Policing by Machine*. Liberty. Available from: https://www.libertyhumanrights.org.uk/issue/policing-by-machine/ (accessed 15 December 2023).

Lum K and Isaac W (2016) To predict and serve? *Significance* 13(5): 14–19.

Mao Y, Wang D, Muller M, et al. (2019) How data scientists work together with domain experts in scientific collaborations: to find the right answer or to ask the right question? *Proceedings of the ACM on Human-Computer Interaction* 3: 1–23.

McKay MH and Richard R (2022) Efficiency of algorithmic policing tools: a nod to C.N. Parkinson. In: The 35th Canadian Conference on Artificial Intelligence, Online, May 2022.

Mittelstadt BD, Allo P, Taddeo M, et al. (2016) The ethics of algorithms: mapping the debate. *Big Data & Society* 3(2): 2053951716679679.

Muir R and O'Connell F (2025) Policing and artificial intelligence. *The Police Foundation*. Available at: https://www.police-foundation.org.uk/publication/policing-and-artificial-intelligence/ (accessed 13 February 2025).

Nichols J, Wire S, Wu X, et al. (2019) Translational criminology and its importance in policing: a review. *Police Practice and Research* 20(6): 537–551.

Nilsson P (2018) *Financial Times*. Available at: https://www.ft.com/content/b34b0b08-ef19-11e8-89c8-d36339d835c0 (accessed 22 September 2023).

Office for Statistics Regulation (2024) The quality of police recorded crime statistics for England and Wales. Available from: https://osr.statisticsauthority.gov.uk/publication/the-quality-of-police-recorded-crime-statistics-for-england-and-wales/pages/3/ (accessed 17 June 2024).

Oswald M, Grace J, Urwin S, et al. (2018) Algorithmic risk assessment policing models: lessons from the Durham HART model and 'experimental' proportionality. *Information and Communications Technology Law* 27(2): 223–250.

O'Connor CD, Ng J, Hill D, et al. (2022) Thinking about police data: analysts' perceptions of data quality in Canadian policing. *The Police Journal: Theory, Practice and Principles* 95(4): 637–656.

Palmiotto F (2021) The black box on trial: the impact of algorithmic opacity on fair trial rights in criminal proceedings. In: Ebers M and Gamito MC (eds) *Algorithmic Governance and Governance of Algorithms*. Springer, pp. 49–70.

Perry WL (2013) *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND.

Police-ML (2024) What is modelling? Available from: https://www.police-ml.com/before-modelling (accessed 25 August 2024).

Prinzel LJ, De Vries H, Freeman FG, et al. (2001) *Examination of automation-induced complacency and individual difference variates*. Technical Memorandum No. TM-2001-211413. Hampton, VA: NASA Langley Research Center.

Richardson R, Schultz JM and Crawford K (2019) Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems, and justice. *NYU Law Review Online* 94: 15. Available from: https://heinonline.org/HOL/LandingPage?handle=hein.journals/nyulro94anddiv=and&iand=&page= (accessed 12 January 2024).

Robinson J (2019) Business change: how forces can use data to drive operational effectiveness and efficiencies. *Policing Insight*. Available from: https://policinginsight.com/feature/opinion/business-change-how-forces-can-use-data-to-drive-operational-effectiveness-and-efficiencies/ (accessed 20 November 2023).

Robinson D and Koepke L (2016) Stuck in a pattern: early evidence on 'predictive policing' and civil rights. *Upturn*. Available from: https://www.upturn.org/static/reports/2016/stuck-in-a-pattern/files/Upturn_-_Stuck_In_a_Pattern_v.1.01.pdf (accessed 19 January 2023).

Sanders C and Henderson S (2013) Police 'empires' and information technologies: uncovering material and organisational barriers to information sharing in Canadian police services. *Policing and Society* 23(2): 243–260.

Saunders J, Hunt P and Hollywood JS (2016) Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology* 12: 347–371.

Sayer H, Polajnar T and Spence R (2024) Algorithmic ambitions: there's no (A)I in team. *Policing Insight*, Available from: https://policinginsight.com/author/hazel-sayer/ (accessed: 24th September 2024).

Schlehahn E, Aichroth P, Mann S, et al. (2015) Benefits and pitfalls of predictive policing. In: *Intelligence and Security Informatics Conference (EISIC)*. IEEE, pp. 145–148.

Selten F, Robeer M and Grimmelikhuijsen S (2022) 'Just like I thought': street-level bureaucrats trust AI recommendations if they confirm their professional judgment. *Public Administration Review* 83(2): 263–278.

Shapiro A (2017) Reform predictive policing. *Nature* 541: 458–460.

Simkute A, Luger E, Jones B, et al. (2021) Explainability for experts: a design framework for making algorithms supporting expert decisions more explainable. *Journal of Responsible Technology* 7: 100017.

Skitka LJ, Mosier KL, Burdick M, et al. (2000) Automation bias and errors: are crews better than individuals? *The International Journal of Aviation Psychology* 10(1): 85–97.

Stanko BA (2007) From academia to policy making: changing police responses to violence against women. *Theoretical Criminology* 11(2): 209–219.

Starr SB (2014) Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review* 66: 803, Available from: https://heinonline.org/HOL/LandingPage?handle=hein.journals/stflr66&div=24&id=&page= (accessed 17 February 2024).

Terpstra J and Kort J (2017) Rigmarole and red tape: background to a common police officers' complaint. *Policing* 11(4): 437–447.

The Law Society (2019). Available from: https://www.lawsociety.org.uk/topics/research/algorithm-use-in-the-criminal-justice-system-report (accessed 4 January 2024).

Urwin S (2017) *Algorithmic forecasting of offender dangerousness for police custody officers: an assessment of accuracy for the durham constabulary model*. Master's Thesis, University of

Cambridge, Cambridge. Available from: https://www.crim.cam.ac.uk/system/files/documents/sheena-urwin-thesis-12-12-2016.pdf (accessed 10 December 2023).
Verma P (2022) The never-ending quest to predict crime using AI. *The Washington Post*. Available from: https://www.washingtonpost.com/technology/2022/07/15/predictive-policing-algorithms-fail/ (accessed 5 January 2024).
Viaene S (2013) Data scientists aren't domain experts. *IEEE XPLORE*. Available from: https://ieeexplore.ieee.org/abstract/document/6674007?casa_token=Grb2PU9KPPAAAAAA:8LtKnfXjQvK5fYjg4NGKmHzfbqffZXysC_Ey_yi54Nfir-xOBB_yD1inmV–w7h_C_D3oW6B (accessed 1 December 2023).
Yeung K (2018) A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. *MSI-AUT* 5. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286027 (accessed 27 January 2024).

Appendix

Acronym

AI	Artificial intelligence
ATRS	Algorithmic transparency recording standard
COMPAS	Correctional offender management profiling for alternative sanctions
HART	Harm assessment risk tool
ML	Machine learning
MLA	Machine learning algorithm
RUDI	Rationale, unification, development, implementation
SSL	Strategic subjects list