

Societal Expectations for AI Deployments in Policing

BAYERL, Petra <<http://orcid.org/0000-0001-6113-9688>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<https://shura.shu.ac.uk/38088/>

This document is the Published Version [VoR]

Citation:

BAYERL, Petra (2026). Societal Expectations for AI Deployments in Policing. In: AKHGAR, Babak and BAYERL, Petra, (eds.) Accountability for Artificial Intelligence in Policing and Security. Security Informatics and Law Enforcement . Cham, Springer, 175-196. [Book Section]

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>



Societal Expectations for AI Deployments in Policing

Petra Saskia Bayerl

10.1 INTRODUCTION

The introduction of artificial intelligence (AI) for policing purposes continues to be a flashpoint in public discussions about both fairness and security. On the one hand, AI is seen as a necessary avenue to improve policing performance (e.g. in the context of community policing, counter-terrorism or cybercrime; [5–7]), and as such has become an integral part of future policing strategies (Chap. 1). On the other hand, legitimate concerns are raised about the potential of AI to make wrong decisions or

This chapter is an adaptation of Bayerl et al. [1].

P. S. Bayerl (✉)
CENTRIC (Centre of Excellence in Terrorism, Resilience, Intelligence and
Organised Crime Research), Sheffield, UK
e-mail: p.s.bayerl@shu.ac.uk

ingrain historical biases. Extreme opinions even consider AI to be ‘unfit for purpose’ [3].

Still, AI deployments by police forces are ongoing, and it is therefore not a question of ‘whether’ but rather of ‘how’ to deploy AI in a way that is acceptable to the public. Yet, public perceptions of AI are not monolithic. Past research shows that public perceptions are split along varying viewpoints from acceptance to resistance (e.g. [2]), ranging from a perceived requirement for the better protection of society to the willingness to destroy surveillance infrastructure to safeguard one’s privacy. Acceptance is further affected by the way that AI is used, such as levels of transparency, the non-sharing of data (i.e. privacy) and the presence of an appeals system [4]. The latter especially speaks to the importance of accountability for public acceptance of AI. Understanding variations in how the public perceives AI deployments in policing is therefore an important aspect to guide AI deployments and to shape AI policies in an effective manner.

This chapter offers insights from a large-scale national citizen consultation in the United Kingdom. It provides a detailed view not only about overall levels of acceptance but also where citizens see the most benefits and concerns in the AI use by police, as well as conditions that would improve acceptance and trust. The subsequent section outlines the survey methodology, followed by a detailed description of the findings. The chapter ends with recommendations for measures to increase public acceptance and trust, based on the evidence of the consultation.

10.2 METHODOLOGY OF THE CONSULTATION

10.2.1 *Data Collection and Participants*

The consultation was conducted as nation-wide anonymous online survey in the UK between April and June 2025, focusing on the general UK population starting from 14 years of age. The recruitment was conducted through an online panel provider and ensured representative sample demographics for gender and age (for age groups 18+) across all four nations in the United Kingdom.

In total, the consultation includes a total of 10,114 participants. Table 10.1 provides an overview of the sample characteristics. As the

overview indicates, the consultation reached a diverse group of participants both in terms of demographics and experiences, which allows the investigation of subgroups within the sample.

10.2.2 Aspects Assessed in the Consultation

The consultation covered seven themes to investigate different aspects of public perceptions and acceptance:

1. Acceptance of, and trust in, AI and AI-supported policing
2. Risks and benefits of AI for personal, societal and institutional policing purposes
3. Measures to improve public trust in the police use of AI
4. The role of accountability for trust
5. Public reactions to 32 deployment scenarios, including current and potential (future) deployment areas
6. Measures to recover trust in case of AI failures (open answers)
7. Trust in the police and other institutions

This chapter focuses on findings with respect to acceptance, risks, benefits and the role of AI accountability.

10.2.3 Ethics

The study received ethics approval by the Sheffield Hallam University's Research Ethics Committee. The survey was conducted as an anonymous questionnaire (i.e. did not collect participant information such as location or IP address). Participants were further required to give informed consent before starting the survey (individuals who did not consent were not allowed to proceed to the survey); double consent was used for the youth sample, requiring parent consent before youths were able to participate in the consultation.

Table 10.1 Sample characteristics

<i>Category</i>	<i>Percentage (%)</i>
<i>Gender</i>	
Women	51.9
Men	47.5
Non-binary/other	0.5
Prefer not to say	0.2
<i>Age</i>	
14–17 years	5.7
18–24 years	19.0
25–34 years	16.6
35–44 years	15.4
45–54 years	16.0
55–64 years	14.4
65–74 years	9.4
75+ years	3.5
Prefer not to say	0.1
<i>Ethnicity</i>	
White	80.1
Black	9.8
Asian	5.2
Mixed ethnicity	3.1
Other	1.5
Prefer not to say	0.3
<i>Education level</i>	
No formal education	0.6
Secondary education	31.7
Bachelor/master degree	34.4
Professional degree	7.9
PhD	4.7
<i>Member of a group discriminated against by police</i>	
No	78.9
Yes	11.2
Unsure	6.6
Prefer not to say	3.3
<i>AI knowledge (general)</i>	
No or very little knowledge	32.0
Some or good knowledge	65.0
Expert	3.0
<i>AI knowledge (policing)</i>	
No or very little knowledge	73.2
Some or good knowledge	25.4

(continued)

Table 10.1 (continued)

<i>Category</i>	<i>Percentage (%)</i>
Expert	1.3
<i>Professional background</i>	
Working in security-related sector (adult sample)	6.3
Parent working in security-related sector (youth sample)	16.2

10.3 FINDINGS

10.3.1 *Extent of Public Support for AI Use by Police*

The consultation demonstrates a largely positive view about AI use by police: 61.3% of participants indicated support for the police use of AI; 18.9% of this group even voiced strong support. A similar percentage (64.0%) moreover considered AI as a capability that enhances policing (cp. Fig. 10.1). The reasons for support can be gathered from the free text comments provided at the end of the survey. These illustrate that support is founded in the expectation that AI makes current policing efforts more effective, as well as the expectation that AI will help the police react more effectively against future crime challenges. AI is thus seen as an important measure to ‘futureproof’ policing capabilities.

The police use of AI was opposed by 19.1% (7.2% of these indicating strong opposition), while 12.4% considered AI as a capability that undermines policing. Reviewing the free comments of these participants, the opposition was grounded in a general lack of trust in the police, as well as fears that ‘the police start seeing [individuals] as data rather than people’. Moreover, AI was seen as ‘too premature’ to be used for policing and security purposes.

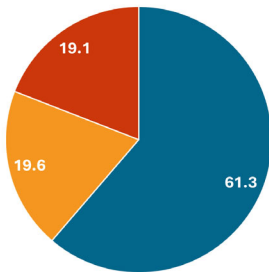
Notable is the considerable number of people with a neutral stance (i.e. neither supportive nor opposing position). The neutral stance was similar in size to the group opposing AI use by police: 19.6% indicated a neutral position compared to 19.1% opposing AI.

An even larger group (23.6%) doubted whether AI has any effect (either positive or negative) for policing. Together, these observations indicate that a considerable majority of the public (80.9%) is not directly opposed to the police use of AI, although some may doubt whether it has any direct effects.

General acceptance of AI

A majority of citizens support AI use by police
% of people saying they want police to use AI

■ Strongly/somewhat support
■ Neutral
■ Strongly/somewhat oppose



A majority of citizens say it enhances policing
% of people saying AI use by police is beneficial

■ Enhances policing
■ Does nothing for policing
■ Undermines policing

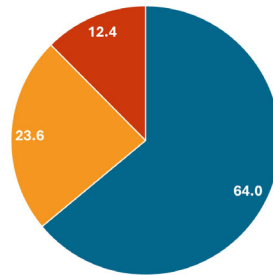


Fig. 10.1 General acceptance of police use of AI (adapted from: Bayerl et al. [1])

Subgroup Differences: investigating subgroups within the sample identified relevant variations. For example, survey participants with higher levels of AI expertise indicated higher levels of acceptance: 88.1% of experts agreed that AI enhances policing, compared to 49.9% of people claiming no expertise in AI. Reversely, 33.9% of people without expertise believed that AI does nothing for policing, compared to only 6.7% of experts. Equally, participants with higher levels of education (PhD and professional degree) showed stronger acceptance: 77.8% and 77.1%, respectively, felt that AI enhances policing compared to 56.9% with secondary school as their highest attainment. While the exact links deserve further investigation, these observations indicate that knowledge and education attainment affect reactions to AI, potentially through greater confidence about the working and effects of AI.

Less clear are results for age groups and gender. Comparing answers in the youth and adult sample suggests that youths are somewhat more critical towards AI, with 59.8% agreeing that AI enhances policing and 26.2% seeing no effect, compared to the adult sample with 64.7% and 23.2%, respectively. Men were slightly more positive compared to women (men: 68.9% enhances policing, 20.0% does nothing versus women: 59.8% and 26.8%, respectively). A clearer effect was found for people identifying as

non-binary with only 30.4% considering AI as enhancing policing (vs 32.6% undermining it). However, this group is extremely small ($n = 46$), which means that these numbers should only be seen as a potential indication for further investigation.

10.3.2 *Main Concerns and Benefits*

To understand public views on concerns and benefits, we differentiated between three areas of impact:

1. **Personal aspects** addressing individual safety, privacy and the safety of the people someone cares about
2. **Societal aspects** addressing effects on communities or public goods such as fundamental rights and effective use of taxes
3. **Institutional aspects** addressing impacts on the ability of the police to fulfil their role

10.3.2.1 *Concerns*

The strongest public concern was the fear that AI could be used to escape the responsibility for failures, i.e. that ‘police pass the blame to AI if something goes wrong’ (cp. Fig. 10.2). This is followed by worries about errors and biases that will negatively affect specific groups. Less concerns were voiced about personal risks, such as individual privacy offline and errors and biases affecting oneself. The most pertinent public concerns therefore seem to be societal and institutional, rather than personal. A relevant observation is that concerns are affected by personal experiences. Perceived vulnerability to police discrimination affected concerns about the police passing blame to AI: 61.1% in this group were concerned or extremely concerned, compared to 48.7% of people who do not see themselves as vulnerable to police discrimination. Crime victimisation in the last 12 months also led to higher rates of concerns: 60.3% in this group voiced concerns about errors and biases that negatively affect specific groups compared to only 36.7% without crime experiences.

10.3.2.2 *Benefits*

The strongest benefits of AI were perceived for enhancing the effectiveness of the police to solve crimes and in an improved ability and more time for the police to deal with serious crimes. 67.8% and 66.4%, respectively,

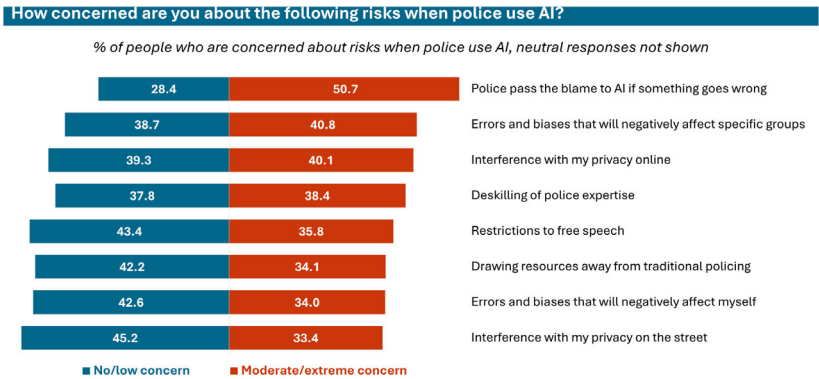


Fig. 10.2 Public concerns against police use of AI (adapted from: Bayerl et al. [1])

of participants indicated that AI makes these aspects better, compared to 12.0% and 11.3%, respectively, who indicated that AI would make them worse (cp. Fig. 10.3). Fewer benefits were seen with regard to the protection of their own safety and even less for the protection of fundamental rights: 47.0% and 37.2%, respectively, indicated that AI makes these aspects better, compared to 17.6% and 28.5%, who indicated that AI makes them worse. This suggests that expected AI benefits are primarily associated with core policing functions, rather than societal and personal benefits.

Similar to overall support, men were more positive across all benefit types: for instance, 28.8% of men indicated that AI makes policing efforts for national security ‘much better’ compared to 17.5% of women. While these are no dramatic differences, this pattern holds across all benefits investigated in the survey.

An interesting effect emerged with people who reported knowledge about AI in policing. Here, benefit expectations were considerably lower compared to people who reported no knowledge of AI in policing. For instance, people without knowledge agreed to 53.7% that AI will improve the ability to protect the safety of people they care about, compared to only 7.7% in the expert group. Again, this pattern holds across all benefits, suggesting a much more critical stance in people with policing AI expertise. This is despite the fact that the overall support for police of AI is higher in police AI experts (63.4% strongly support) than in lay people (10.4% strongly support).

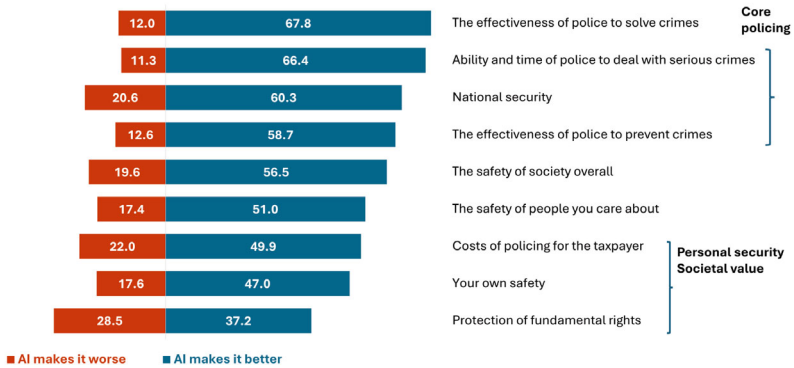
In your view, how does the police use of AI affect...*% of people who perceive benefits when police use AI; answers for 'no effect' not shown*

Fig. 10.3 Benefits perceived from the use of AI by police (adapted from: Bayerl et al. [1])

This imbalance is a strong indication that how support is measured has a considerable impact on results and may tap into different aspects of public attitudes, and that overall acceptance does not guarantee acceptance for specific usages or benefits narratives.

10.3.2.3 No Effect of AI for Policing

Additional insights come from answers that claim that ‘AI has no effect’ on policing. More than a third of the public considered AI as being ineffective for their own safety (35.4%), the safety of people they care about (31.5%), as well as the protection of fundamental rights (31.5%). Fewer doubts existed on the impact of AI for the effectiveness of police to solve crimes (20.2% indicate ‘no effect’) and on national security (19.2% indicate ‘no effect’, cp. Fig. 10.4). Similar to overall benefits, these findings underscore that the public sees AI’s main effect in core policing functions, rather than personal ones.

10.3.3 Public Trust in AI Use by Police and AI Decisions

The picture on public trust is mixed. Slightly over 50% indicated that they trust the police to use AI to support their decision-making, 15.4% of these ‘very much’ (cp. Fig. 10.5). This still leaves 1 in 4 (25.0%) who

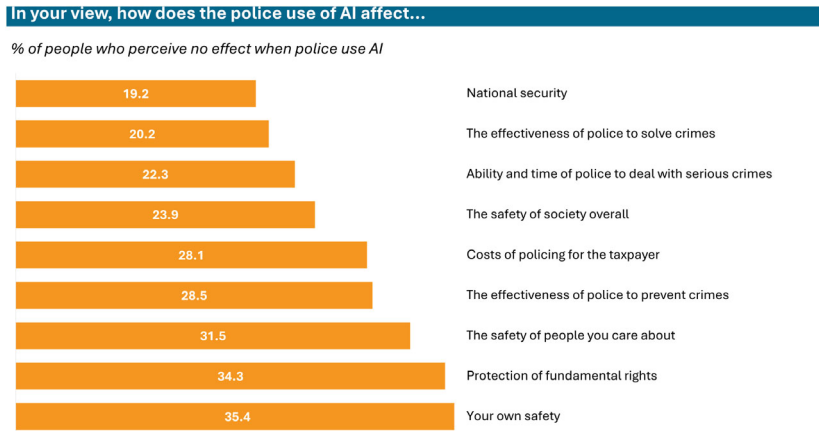


Fig. 10.4 Perception of ‘no benefits’ from AI use by police across deployment areas (adapted from: Bayerl et al. [1])

distrusted AI for decision support and 1 in 5 (20.2%) who were neutral about the issue. Public trust was considerably lower when AI is allowed to make its own decisions for policing tasks: only 36% trusted AI to make its own decisions (9.4% of these ‘very much’), compared to 39.7% who distrust AI-driven decisions (15.5% ‘very much’; 24.3% ‘neutral’).

As an additional measure of trust, we further asked how willing someone would be to provide their own personal information to train AI for policing and security purposes. The question reveals a high level of unwillingness or at least uncertainty: 40% would be (very) unwilling to provide their personal information for AI training, while 29% were unsure (cp. Fig. 10.6). This leaves a very small group of 9.3% who would be ‘very

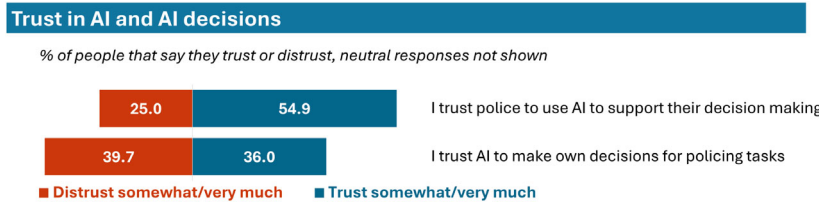


Fig. 10.5 Trust in police using AI (adapted from: Bayerl et al. [1])

Willingness to provide own data

% of people that say they are willing to provide own data

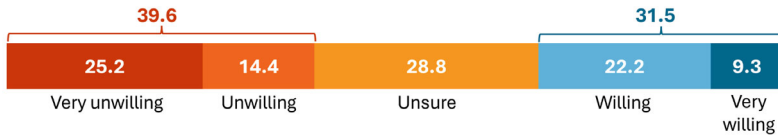


Fig. 10.6 Willingness to provide personal data for training of policing AI systems (adapted from: Bayerl et al. [1])

willing' to provide their information. Working in a security-related profession increased the willingness to provide data to some extent: 43.9% in this group would be willing to provide their own data for the training of police AI (15.5% of these very willing), compared to 31.3% who are not linked to a security-related profession (with 9.2% of these very willing).

10.3.4 Public Reactions to Different AI Deployment Scenarios

As the section on specific benefits demonstrates (Sect. 10.3.2.2), public attitudes are impacted by the specific purpose for which the police deploy their AI capabilities. As part of the consultation, we therefore included 32 disparate AI deployment scenarios to map out systematic variations in public reactions.

The 32 scenarios covered existing as well as possible (future) AI uses in policing and security. The purposes varied with respect to their deployment area (e.g. migration, child sexual exploitation, counter-terrorism, burglary, forensics, etc.), the deploying organisation (police, UK Home Office, UK Border Forces), their purpose (crime prevention, crime investigation including the identification of perpetrators, internal police efficiencies, etc.) and the risk level for privacy intrusions (low vs. high). The full list of AI deployment scenarios is provided in Table 10.2.

For each deployment scenario we assessed:

- its general acceptability,
- the degree to which AI is trusted in this context,
- the expected level of monitoring by an independent oversight body.

Table 10.2 List of AI deployment scenarios

Scenario 1: Reviewing evidence and producing a case file are traditionally done manually by police officers. Your local police force uses an AI tool to automatically review the evidence and to produce a case file summary. The AI-generated case summaries are used by solicitors to decide whether to charge someone with a crime. Scenario 2: To prevent small boat crossings from France to UK, UK border forces use AI sensors to flag up potential crossings in France. The sensors automatically notify UK border forces to intercept the boats and prevent them from landing in the UK. Scenario 3: UK border police uses AI to investigate the online footprint of all foreign citizens who come through British borders. The automatic scan flags up any indications of criminal/terrorist activities or communication online to identify potential hostile intent against UK and impact decisions whether a person can enter the country. Scenario 4: Your local police force uses an AI-driven robot community policing officer to autonomously patrol your neighbourhood and areas associated with high crime. If the robot identifies a potentially critical situation, it will contact the closest police station or other emergency services. Scenario 5: Your local police is aware that a child is at risk of sexual exploitation. A fast-food restaurant handed over CCTV pictures of a meeting between the child and an unknown adult. The police places smart cameras in the area that scans people's faces to find the adult involved. Scenario 6: Investigating child sexual abuse material causes psychological stress and mental health issues in many officers. Your local police force deploys an AI tool to scan images/videos of child sexual abuse. The AI tool automatically assigns grades of severity of the abuse safeguarding officers from viewing the material. Scenario 7: AI tools are increasingly used to generate large quantities of child sexual abuse material. This poses a problem for police, as they do not know whether the children, they are trying to identify, are real or AI generated. Your local police introduce an AI tool to identify whether an image/video is AI generated, so police can focus their efforts on helping real victims. Scenario 8: Your local police force uses an AI tool that scans social media for signs of impending terrorist attacks in your city. The AI creates alerts for police to start direct surveillance of suspects, or in case of an immediate threat, police receive an automatic request to arrest them. Scenario 9: Your local police force uses an AI tool that scans social media for signs of serious and organised crime. This includes people advertising drugs for sale, evidence of firearms use and human trafficking. AI helps the police to identify the perpetrators and to safeguard victims. Scenario 10: An airport uses an AI trained on gait analysis to identify unusual behaviours, such as nervousness, hesitation or erratic movements that could indicate terrorist intent. Any warning markers are sent to airport security and police to initiate a response. Scenario 11: Your local police force is investigating a burglary. Very little evidence has been recovered, but forensics officers identified a partial fingerprint. The police force uses AI to enhance the print to then search for matches in a national fingerprint database.

(continued)

Table 10.2 (continued)

Scenario 12: a police force uses an AI tool to analyse DNA samples from cold cases. The AI tool uses predictive analysis to reconstruct missing sequences of partial DNA profiles, which the police force uses to search for matches in a national DNA database

Scenario 13: Your local police force uses an AI tool to speed up their investigations into digital devices such as mobile phones. When a phone is seized from a suspect, the AI searches the phone to quickly assess location data, images, records of phone calls and the content of text communication that may be relevant to the case

Scenario 14: Your local police force uses an AI tool to scan online profiles of local groups for evidence of hate speech (e.g. violence against vulnerable communities or people based on race, ethnicity, gender, religion or sexual orientation). The AI identifies individuals that seem to post hate speech and flags them for further investigation

Scenario 15: AI-generated content of politicians, actors and other celebrities (deepfakes) are being used to spread false information and encouraging public disorder or hate. Your local police force uses AI investigation tools to automatically identify the creators of AI-generated deepfake videos to prevent further spread of disinformation

Scenario 16: The courts use AI to speed up the justice process. The AI decides whether charged suspects should be bailed by the court or be remanded in prison and also calculates sentences for those found guilty

Scenario 17: The Home Office deploys an AI to speed up decisions regarding early release of prisoners. The AI reviews applications made by prisoners and decides, whether they can serve the remainder of their sentence in the community

Scenario 18: The Home Office uses an AI-based visa and immigration system to make decisions more quickly. The AI automatically assesses visa and settlement applications and decides on rejection of unsuitable applicants

Scenario 19: UK security services use AI to detect insider threats to the security of national infrastructure. They specifically aim to protect water and electricity supply as well as nuclear reactions from foreign hostile actors

Scenario 20: Your local police force uses an AI system to automate the purchase of uniforms and similar items to free up police time. The AI system identifies low stock levels, assesses the suitability of an item offered for sale and directly submits a purchase order to the company

Scenario 21: Before a demonstration, your local police use AI capabilities to monitor social media feeds for signs of potential intent of violence or crimes. People who are flagged up are prevented from joining the protest

Scenario 22: During a demonstration, your local police use AI-based sensors to automatically identify locations where unrest may break out and lead to violence. If the sensors detect potentially problematic patterns, the AI system automatically direct police resources to these areas

Scenario 23: Your local police force uses an AI system to take emergency 999 calls. The AI interacts with the caller and decides whether and when to dispatch response officers

(continued)

Table 10.2 (continued)

Scenario 24: a case of shoplifting at a large supermarket in your area has been reported to the police. The stolen items are worth just under £100. The supermarket has handed over pictures of its surveillance camera to your local police force. The police run AI-based facial recognition on the pictures to identify potential suspects
Scenario 25: a disused warehouse in your area is suspected to be used to store both weapons and drugs for an organised crime group. A nearby surveillance camera has captured a facial image of an individual frequently leaving with packages. The police set up an AI-based live facial recognition system to identify the person in real time
Scenario 26: When on the beat, officers in your local police force use smart glasses that access information about people on the street in real time. If the AI flags up concerns about a person, the officers initiate a stop-and-search
Scenario 27: Your local police force uses an AI system to take non-emergency 101 calls. The AI interacts with the caller and provides advice to their queries
Scenario 28: Police training programmes in your local police utilise AI to generate realistic training scenarios. Police students learn through interaction with highly convincing AI-controlled virtual witnesses, victims, or suspects
Scenario 29: Your local police force uses an AI-assisted system while interviewing suspects. The AI analyses words, micro-expressions and tone of voice of suspects to assess credibility and to provide investigators with guidance for questioning
Scenario 30: Your local police force uses an AI system to scan social media profiles of police officers. The AI automatically flags up officers potentially acting disproportionately or exhibiting racist, discriminatory, or otherwise problematic behaviour
Scenario 31: Your local police force uses AI to analyse police records, social media and publicly available data for indications of domestic violence to recommend preventive interventions for potential victims
Scenario 32: Your local police force deploys automated, AI-controlled police vehicles that autonomously patrol areas and independently identify and report suspicious activities, traffic violations and accidents

10.3.4.1 Acceptance Levels across Usage Situations

The comparison across scenarios demonstrates how strongly public reactions are influenced by the type of AI deployment situation. The lowest levels of acceptance were found for AI-automated 999 emergency calls (scenario 23) and automated decisions about the early release of criminal suspects and prisoners (scenario 16) with only 27.4% and 29.3% in the sample indicating some or strong acceptance. In contrast, the highest levels of acceptance were observed for AI use to identify (potential) perpetrators of child sexual exploitation (62.4%; scenario 5), perpetrators of shoplifting in a neighbourhood store (58.7%; scenario 24) and (potential) terrorists crossing UK borders (58.0%; scenario 3). These findings

suggest that the identification of potential suspects is one of the purposes where AI finds most acceptance.

10.3.4.2 Trust Levels across Usage Situations

Considerable variation in opinions can also be observed for the level of trust in AI use by policing. Again, trust was lowest where AI may cause direct harms to individuals themselves, for instance, where AI replaces 999 dispatchers (26.3%; scenario 23) or where AI makes decisions about an early release of prisoners (26.6%; scenario 16). Scenarios with low trust and low acceptance levels were thus closely linked.

In contrast, the highest degree of public trust emerged for using AI to improve the administrative efficiency of police forces by automating the purchase of police uniforms (54.5%; scenario 20) but also again for the protection of a child from sexual exploitation (53.2%; scenario 5). This seems to indicate that trust in police use of AI is highest both for situations in which AI deployments carry low risks for the public and for situations in which AI can reduce serious harms to vulnerable groups. It is worth noting, however, that even the highest trust levels are only slightly above 50%, indicating a limited degree of trust in AI use by police overall for specific usage scenarios.

10.3.4.3 Need for Independent Oversight

The context also affects the degree to which the public expects monitoring by an independent body for each AI deployment. The strongest need for independent oversight emerged for scenarios in which AI makes its own decisions in high-risk situations (e.g. when AI is used to automatically review evidence (63.7%; scenario 1) or to automatically identify AI-generated child sexual exploitation material (62.1%; scenario 7)). The lowest need for independent oversight was found for the purchase of uniforms (36.7%; scenario 20) and the deployment of AI sensors during demonstrations to direct police resources (45.2%; scenario 22).

10.3.4.4 Links Between Acceptance, Trust and Independent Oversight

When comparing acceptance, trust and need for oversight, it is clear that oversight by independent bodies is a broadly shared concern, independent of acceptance and trust.

As may be expected, results show strong expectations for independent oversight in situations with low trust and acceptance (e.g. scenario 23 in

which an AI system takes emergency 999 calls or scenario 17 in which AI is deployed to speed up decisions regarding the early release of prisoners). However, independent oversight is also expected in scenarios that show considerable levels of trust and acceptance (e.g. scenario 7 in which an AI tool identifies whether a CSAM image/video is AI generated). Monitoring by an independent oversight body is thus not only needed in situations of low public acceptance or trust but expected by the public whenever police deploy AI.

10.3.5 *The Role of AI Accountability for Public Trust*

The importance of independent oversight together with concerns about negative AI consequences and of police ‘passing blame to AI if something goes wrong’ points to questions of accountability and trust in policing organisations.

Asked directly, how much they would trust police to use AI accountably, slightly over 50% indicated they would (16.3% of these very much). This leaves 1 in 5 (21.1%) who were neutral on the issue and 1 in 4 (25.6%) who indicated distrust in police to use AI accountably.

To get a better understanding of how to improve trust in AI policing, we provided five measures that are commonly proposed as safeguards (see Fig. 10.7). *Explainability* and *justification* emerged as the least effective measure from the public’s perspective: 39.0% of respondents signalled that this would give them only slight trust or no trust at all (cp. Fig. 10.7). This is higher than the 30.4% who indicated that this measure would give them considerable or even very high trust. *Compensation and redress* fared slightly better (33.3% considerable/very high trust vs. 35.5% limited/no trust), as did enforcement of *changes to the AI system* in cases where the police use of AI leads to negative consequences (33.4% considerable/very high trust vs. 35.2% limited/no trust). The most effective measures, and the only ones where ‘trust’ percentages were higher than those for ‘no trust’, were the close monitoring of AI use and effective accountability (38.5% and 40.0%, respectively). Thus, none of the five common trust measures may be able to instil trust by themselves; however, the results do give indications to public preferences for assurance measures.

In this context it is compelling that 75.5% of the public wanted a mandatory accountability process before the police deploy AI. However, less than 20% of the public thought that AI accountability in policing is sufficiently established at this moment (cp. Fig. 10.8).

How would the following measures affect your trust in police use of AI?

% of people that say this measure give them trust in police use of AI

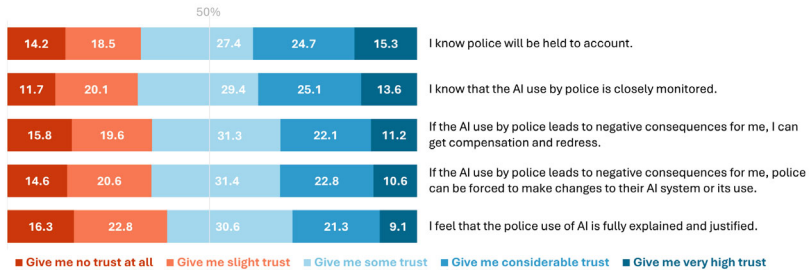


Fig. 10.7 Degree of trust due to different accountability measures (adapted from: Bayerl et al. [1])

AI Accountability

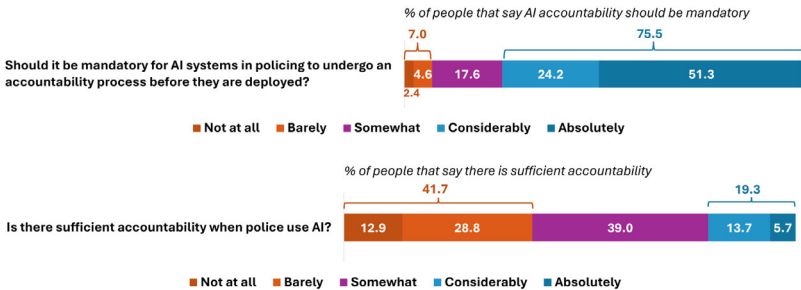


Fig. 10.8 Perceptions of current and required AI accountability (adapted from: Bayerl et al. [1])

10.4 IMPLICATIONS FOR AI DEPLOYMENTS

Results overall show that a majority of the UK public supports AI in policing, particularly where it enhances the effectiveness of policing and addresses serious harms such as child sexual exploitation and terrorism. In the same regard, findings suggest that AI should be used to support (human) decision-making, as trust drops sharply when AI is expected to operate autonomously in making policing decisions. This may be explained by the fact public concerns centre around accountability gaps, such as a fear that police ‘will blame AI’ if decisions go wrong and the view that accountability is currently (somewhat to totally) lacking.

Accordingly, there is a widely shared expectation of independent oversight, even for deployments with high levels of acceptance. Independent oversight thus must be universal—with an expectation of mandatory accountability checks before AI deployments.

The consultation thus gives important guidance on citizen expectations on how to approach and implement AI use in policing. Policing should, as a minimum:

- Establish statutory, independent oversight and accountability mechanisms.
- Enforce mandatory pre-deployment reviews and bias testing.
- Strengthen transparency and citizen engagement to maintain legitimacy.
- Apply a tiered risk-based deployment model, prioritising crime-solving and child protection, while restricting high-risk uses such as AI-led 999 triage and prisoner release.

Together, these measures will create an ‘accountability-first’ AI model, which can harness AI’s benefits for policing while supporting public acceptance and trust. The following section provides detailed recommendations on how to achieve accountable AI deployments in policing in line with public expectations.

10.4.1 Recommendations for AI Deployments in Policing to Ensure AI Accountability and Support Public Trust

10.4.1.1 Measures for Governance and Oversight

- Create a statutory AI Oversight Authority for policing and security that is independent from policing and law enforcement, with powers to monitor, audit and sanction the improper use of AI.
- Mandate pre-deployment accountability reviews for all AI systems, including tests for fairness, reliability and proportionality.
- Require regular post-deployment audits and publish findings to appropriate authorities and oversight bodies.
- Build cross-institutional coordination to avoid siloed regulation.

10.4.1.2 *Measures for Enshrining Accountability and Responsibility*

- Introduce a national framework, modelled on Ombudsman-style approaches, that regulates redress and compensation when AI errors cause errors or harm.
- Require record-keeping ('AI decision logs') for all uses in critical policing tasks to ensure the traceability for courts, oversight bodies and citizens.
- Caution against the transfer of 'blame' from police to AI systems; ensure accountability always rests with the provider and/or user deploying the AI.

10.4.1.3 *Measures to Improve Transparency and Public Communication*

- Implement Transparency-by-Default requiring police forces to disclose where, why and how AI is being deployed, including details of training data sources (under consideration of operation/legal restrictions such as national security or victim protection).
- Produce plain-language impact assessments for each AI deployment that are available to the public (under consideration of operation/legal restrictions).
- Ensure public awareness of AI deployments by clarifying benefits, risks and limits of AI in policing to counter misinformation.

10.4.1.4 *Measures for Deployment and Risk Management*

- Implement a tiered deployment model:
 - Light oversight for low-risk efficiency AI (e.g. logistics, scheduling, admin).
 - Monitoring with human-in-the-loop for medium-risk operational AI (e.g. crime pattern analysis, forensics support, life facial recognition).
 - Tight restriction or prohibition for high-risk decision AI (e.g. sentencing, prisoner release, emergency call triage).
- Require final human decision-making for any system that affects liberty, safety, rights (e.g. detention, bail, surveillance) or complaints.

- Review AI uses with low public acceptance and high risk (e.g. AI replacing 999 operators, autonomous patrol robots) to ensure benefits outweigh the risks.

10.4.1.5 *Measures to Remove/Mitigate Biases and Ensure Fairness and Equality*

- Require bias testing across demographic groups before AI procurement and during AI deployment.
- Establish documentation of any known biases and, where required, ensure appropriate mitigation measures.
- Establish citizen-informed fairness benchmarks, especially where vulnerable or historically over-policed groups are involved.
- Ensure the independent testing of systems before approval.
- Mandate regular Equality Impact Assessments for AI tools in use.

10.4.1.6 *Measure for Privacy and Data Protection*

- Introduce a ‘necessity and proportionality test’ before deploying AI with large-scale data collection from the public (e.g. facial recognition, social media monitoring).
- Require data minimisation and purpose limitation in AI deployments, aligned with GDPR and relevant national regulations.

10.4.1.7 *Measure for Improved Citizen Engagement and Trust-Building*

- Establish citizen panels or advisory councils on AI policing to provide continuous input into governance.
- Institutionalise regular citizen consultations at national and regional levels to capture evolving public sentiment.
- Ensure citizen consultations comprise a comprehensive spectrum of citizen demographics and capture subgroup-specific benefit expectations, as well as subgroup-specific concerns.
- Incorporate youth perspectives into AI accountability debates, recognising generational differences in trust.

10.4.1.8 *Measures for Capacity Building and Training*

- Provide specialist AI literacy training for officers and staff to ensure they understand system limits and ethical considerations.
- Develop cross-disciplinary training for oversight bodies (law, ethics, data science, human rights) with a specific focus on AI accountability.
- Build a central AI in Policing Knowledge Hub for sharing best practices, red flags and audit results across police forces and connected stakeholders.

Acknowledgements Work in this chapter was partially funded under the AIPAS project ('AI Accountability for Policing and Security'; UKRI Engineering and Physical Sciences Research Council, Responsible AI (RAI) IA091 Grant Ref: EP/Y009800/1).

REFERENCES

1. P.S. Bayerl, B. Akhgar, F. Sampson, M. Snowden, C. Rowley, A. Paradise, Spencer, C, Citizen Consultation on the Police Use of Artificial Intelligence (AI) in the UK: Acceptance, Trust and Accountability. AIPAS Report (2025). Document available online: [https://stwebsitesprodeu.blob.core.windows.net/centric-website/CENTRIC_AI+Policing_National-Citizen-Consultation\(public-version\).Oct2025-WEB.pdf](https://stwebsitesprodeu.blob.core.windows.net/centric-website/CENTRIC_AI+Policing_National-Citizen-Consultation(public-version).Oct2025-WEB.pdf)
2. Y. Ezzeddine, P.S. Bayerl, H. Gibson, Safety, privacy, or both: Evaluating citizens' perspectives around artificial intelligence use by police forces. *Polic. Soc.* **33**(7), 861–876 (2023). <https://doi.org/10.1080/10439463.2023.2211813>
3. W. Heaven, Predictive Policing Algorithms Are Racist. They need to be dismantled (2020). Retrieved from MIT Technology Review: www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice
4. L. Horvath, O. James, S. Banducci, A. Beduschi, Citizens' acceptance of artificial intelligence in public services: evidence from a conjoint experiment about processing permit applications. *Gov. Inf. Q.* **40**(4) (2023). <https://doi.org/10.1016/j.giq.2023.101876>
5. T.W. Hung, C.P. Yen, On the person-based predictive policing of AI. *Ethics Inf. Technol.* **23**, 165–176 (2021)
6. R. Pastor, F. Mignet, T. Mattes, A. Gurzawska, H. Nitsch, D. Wright, COPKIT: technology and knowledge for early warning/early action-led

- policing in fighting organised crime and terrorism, in *Technology Development for Security Practitioners. Security Informatics and Law Enforcement*, ed. by B. Akhgar, D. Kavallieros, E. Sdongos (Springer, Cham, 2021), pp. 121–133
7. J. Seaman, Cyber threat prediction and modelling, in *Artificial Intelligence and National Security*, ed. by R. Montasari (Springer, Cham, 2022), pp. 113–156

Petra Saskia Bayerl is Professor of Digital Communication and Security and Head of Research at CENTRIC. Her research addresses societal implications and governance of new technologies in security and policing. She has been co-lead for both the AP4AI and AIPAS projects and lead for national and EU-projects to improve the protection against online harms.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits any noncommercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if you modified the licensed material. You do not have permission under this license to share adapted material derived from this chapter or parts of it.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

